Advances in Cancer Screening

Cancer Treatment and Research

Steven T. Rosen, M.D., *Series Editor*

Nathanson, L (ed): Malignant Melanoma: Biology, Diagnosis, and Therapy. 1988. ISBN 0-89838-384-6
Pinedo HM, Verweij J (eds): Treatment of Soft Tissue Sarcomas. 1989. ISBN 0-89838-391-9
Hansen HH (ed): Basic and Clinical Concepts of Lung Cancer. 1989. ISBN 0-7923-0153-6
Lepor H, Ratliff, TL (eds): Urologic Oncology. 1989. ISBN 0-7923-0161-7
Benz C, Liu, E (eds): Oncogenes. 1989. ISBN-7923-0237-0
Ozols RF (ed): Drug Resistance in Cancer Therapy. 1989. ISBN 0-7923-0244-3
Surwit EA, Alberts DS (eds): Endometrial Cancer. 1989. ISBN 0-7923-0286-9
Champlin R (ed): Bone Marrow Transplantation. 1990. ISBN 0-7923-0612-0
Goldenberg D (ed): Cancer Imaging with Radiolabeled Antibodies. 1990. ISBN 0-7923-0631-7
Jacobs C (ed): Carcinomas of the Head and Neck. 1990. ISBN 0-7923-0668-6
Lippman ME, Dickson R (eds): Regulatory Mechanisms in Breast Cancer: Advances in Cellular and Molecular Biology of Breast Cancer, 1990. ISBN 0-7923-0868-9
Nathanson L (ed): Malignant Melanoma: Genetics, Growth Factors, Metastases, and Antigens. 1991. ISBN 0-7923-0895-6
Sugarbaker PH (ed): Management of Gastric Cancer. 1991. ISBN 0-7923-1102-7
Pinedo HM, Verweij J, Suit HD, (eds): Soft Tissue Sarcomas: New Developments in the Multidisciplinary Approach to Treatment. 1991. ISBN 0-7923-1139-6
Ozols RF (ed): Molecular and Clinical Advances in Anticancer Drug Resistance. 1991. ISBN 0-7923-1212-0
Muggia FM (ed): New Drugs, Concepts and Results in Cancer Chemotherapy. 1991. ISBN 0-7923-1253-8
Dickson RB, Lippman, M.E. (eds): Genes, Oncogenes and Hormones: Advances in Cellular and Molecular Biology of Breast Cancer. 1992. ISBN 0-7923-1748-3
Humphrey G Bennett, Schraffordt Koops H, Molenaar WM, Postma, A. (eds): Osteosarcoma in Adolescents and Young Adults: New Developments and Controversies. 1993. ISBN 0-7923-1905-2
Benz CC, Liu ET (eds): Oncogenes and Tumor Suppressor Genes in Human Malignancies. 1993. ISBN 0-7923-1960-5
Freireich EJ, Kantarjian H (eds): Leukemia: Advances in Research and Treatment. 1993. ISBN 0-7923-1967-2
Dana BW (ed): Malignant Lymphomas, Including Hodgkin's Disease: Diagnosis, Management, and Special Problems. 1993. ISBN 0-7923-2171-5
Nathanson L (ed): Current Research and Clinical Management of Melanoma. 1993. ISBN 0-7923-2152-9
Verweij J, Pinedo HM, Suit HD (eds): Multidisciplinary Treatment of Soft Tissue Sarcomas. 1993. ISBN 0-7923-2183-9
Rosen ST, Kuzel TM (eds) Immunoconjugate Therapy of Hematologic Malignancies. 1993. ISBN 0-7923-2270-3
Sugarbaker PH (ed): Hepatobiliary Cancer. 1994. ISBN 0-7923-2501-X
Rothenberg ML (ed): Gynecologic Oncology: Controversies and New Developments. 1994. ISBN 0-7923-2634-2
Dickson RB, Lippman ME (eds.): Mammary Tumorigenesis and Malignant Progression. 1994. ISBN 0-7923-2647-4
Hansen HH (ed): Lung Cancer. Advances in Basic and Clinical Research. 1994. ISBN 0-7923-2835-3
Goldstein LJ, Ozols RF (eds.): Anticancer Drug Resistance. Advances in Molecular and Clinical Research. 1994. ISBN 0-7923-2836-1
Hong WK, Weber RS (eds.): Head and Neck Cancer. Basic and Clinical Aspects. 1994. ISBN 0-7923-3015-3
Thall PF (ed): Recent Advances in Clinical Trial Design and Analysis. 1995. ISBN 0-7923-3235-0
Buckner CD (ed): Technical and Biological Components of Marrow Transplantation. 1995. ISBN 0-7923-3394-2
Muggia FM (ed): Concepts, Mechanisms, and New Targets for Chemotherapy. 1995. ISBN 0-7923-3525-2
Klastersky J (ed): Infectious Complications of Cancer. 1995. ISBN 0-7923-3598-8
Kurzrock R, Talpaz M (eds): Cytokines: Interleukins and Their Receptors. 1995. ISBN 0-7923-3636-4
Sugarbaker P (ed): Peritoneal Carcinomatosis: Drugs and Diseases. 1995. ISBN 0-7923-3726-3
Sugarbaker P (ed): Peritoneal Carcinomatosis: Principles of Management. 1995. ISBN 0-7923-3727-1
Dickson RB, Lippman ME (eds.): Mammary Tumor Cell Cycle. Differentiation and Metastasis. 1995. ISBN 0-7923-3905-3
Freireich EJ, Kantarjian H (eds.): Molecular Genetics and Therapy of Leukemia. 1995. ISBN 0-7923-3912-6
Cabanillas F, Rodriguez MA (eds.): Advances in Lymphoma Research. 1996. ISBN 0-7923-3929-0

Advances in Cancer Screening

edited by

ANTHONY B. MILLER
University of Toronto, Canada

1996 **KLUWER ACADEMIC PUBLISHERS**
BOSTON / DORDRECHT / LONDON

Distributors for North America:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061 USA

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS

Library of Congress Cataloging-in-Publication Data
Advances in cancer screening / edited by Anthony B. Miller.
p. cm. — (Cancer treatment and research; 86)
Includes bibliographical references and index.
ISBN-13: 978-1-4612-8539-7 e-ISBN-13: 978-1-4613-1265-9
DOI: 10.1007/978-1-4613-1265-9
1. Cancer — Diagnosis. 2. Cancer — Prevention. 3. Medical screening. I. Miller, A. B. (Anthony B.) II. Series: Cancer treatment and research; v. 86.
[DNLM: 1. Neoplasms — prevention & control. 2. Mass Screening — methods. W1 CA693 v. 86 1996 / QZ 200 A2442 1996]
RA645.C3A34 1996
362.1'96994 — dc20
DNLM/DLC
for Library of Congress 96-7537
CIP

Contents

Contributors

MARK L., Bernstein, M.D., Montreal Children's Hospital, 2300 Tupper Street, Montreal, Quebec, H3H 1P3, Canada

NICHOLAS E., Day, Ph.D., MRC Biostatistics Unit, The Institute of Public Health, University of Cambridge, University Forvie Site, Robinson Way, Cambridge CB2 2SR, United Kingdom

MARK J., Elwood, M.D., FRCP(C), Director, Hugh Adam Cancer Epidemiology Unit, Faculty of Preventive and Social Medicine, University of Dunedin, PO Box 913, Dunedin, New Zealand

WILLIAM D., Foulkes, Ph.D., MRCP(UK), Division of Medical Genetics, Department of Medicine, McGill University, 1650 Cedar Avenue, Montreal, Quebec, H3G 1A4, Canada

JOHN K., Gohagen, Ph.D., Biometry Branch DCPC, National Cancer Institute, Executive Plaza North, Suite 344, 6130 Executive Blvd. MSC 7354, Bethesda, MD 20892-7392, USA

MATTI Hakama, Sc.D., University of Tampere, Department of Public Health, Box 607 SF-33101 Tampere, Finland

MURRAY Krahn, M.D., FRCP(C), Division of General Internal Medicine and Clinical Epidemiology, The Toronto Hospital, Eaton Wing N, 200 Elizabeth Street, Ground Floor, Rm 248, Toronto, Ontario, M5G 2C4, Canada

BARNETT S., Kramer, M.D., Biometry Branch DCPC, National Cancer Institute, Executive Plaza North, Suite 344, 6130 Executive Blvd. MSC 7354, Bethesda, MD 20892-7394, USA

JACK S., Mandel, Ph.D., MPH, School of Public Health–Environmental & Occupational Health, University of Minnesota, Box 807 UMHC, 420 Delaware SE, Minneapolis, MN 55455, USA

ANTHONY B., Miller, M.B., FRCP, Department of Preventive Medicine and Biostatistics, University of Toronto, 12 Queens Park Crescent West, Toronto, Ontario, M5S 1A8, Canada

SUE M., Moss, Ph.D., Cancer Screening Evaluation Unit, The Institute of Cancer Research, Royal Cancer Hospital, Block D, Cotswold Road, Sutton, Surry, SM2 5NG United Kingdom

GARY Naglie, M.D., FRCP(C), The Toronto Hospital, 238-200 Elizabeth Street, Toronto, Ontario, M5G 2C4, Canada

STEVEN A., Narod, M.D., FRCP(C), University of Toronto/Women's College Hospital, Department of Breast Cancer Research, Women's College Hospital, 76 Grenville Street, Toronto, Ontario, M5S 1B2, Canada

MAXWELL D., Parkin, Ph.D., Unit of Descriptive Epidemiology, International Agency for Research on Cancer, 150, Cours Albert Thomas, 69372 Lyon, Cedex 08 France

PAOLA Pisani, Ph.D., Unit of Descriptive Epidemiology, International Agency for Research on Cancer, 150, Cours Albert Thomas, 69372 Lyon, Cedex 08 France

ARNOLD L., Potosky, Ph.D., Biometry Branch DCPC, National Cancer Institute, Executive Plaza North, Suite 344, 6130 Executive Blvd. MSC 7354, Bethesda, MD 20892-7394, USA

PHILIP C., Prorok, Ph.D., Biometry Branch DCPC, National Cancer Institute, Executive Plaza North, Suite 344, 6130 Executive Blvd. MSC 7354, Bethesda, MD 2892-7394, USA

WILLIAM G., Woods, M.D., Department of Pediatrics, Box 454, University of Minnesota, 420 Delaware Street, Minneapolis, MN 55455, USA

Preface

Screening for cancer is an important focus of cancer control. Yet screening, since it involves administering a test to large segments of the population deemed at risk for the disease of interest, is potentially a major consumer of scarce health care resources. Further, the benefits sought from cancer screening, particularly reduction in mortality from the disease, are not always realized, sometimes for biological reasons and sometimes for organizational ones. Thus the paradigm that "early detection must always be beneficial," taught to health care professionals and publicized widely through the media to the public, has been challenged in the last two decades for a number of cancer sites. It is the purpose of this volume to determine the extent to which the requirements for the introduction of population-based screening programs have been met, as a result of extensive research on screening during the last two decades, with the major concentration on findings from the recent decade.

The volume addresses important issues for the majority of the sites for which data on the effectiveness of screening are currently available. It also addresses some general principles that apply to screening and pays attention to the advances in understanding of the genetic basis of some cancers, which are beginning to raise important ethical as well as practical issues. The viewpoint is largely that of epidemiology, and many of the authors are former contributors to the series of publications on evaluation of screening for cancer that arose from the program of the International Union Against Cancer Project on Screening for Cancer, which I had the honor to chair for many years [1–5]. Thus the authors of the various chapters have carefully evaluated the available evidence on the effectiveness of screening and have usually been able to reach a conclusion as to whether screening for the relevant cancer site should be part of a cancer-control, public health policy approach. However, clinical aspects have not been neglected, especially when they form a major part of the approach to screening for a particular site (as for colorectal cancer and melanoma screening, for example), or where part of the concern has to be whether adequate treatment is available for the abnormalities identified by screening (as for screening for cancer in high-risk families). Indeed, we must constantly be aware of the requirement for effective screening that there should be effective treatment — a potential difficulty of management for many of the

lesions found after screening for cervical cancer, and also for breast cancer, perhaps particularly in women under the age of 50.

This volume has been planned as a valuable resource document for all interested in cancer screening, including government and nongovernment organizations concerned with cancer control; cancer researchers; and members of national cancer societies and of international organizations concerned with cancer.

I should like to pay tribute to the many authors who have contributed to this volume, and who, with considerable patience, acceded to my urgings over various matters of detail. This has truly been a multi-authored production, though with a great deal of mutual understanding of objective and purpose.

A.B. Miller
Toronto, Canada

References

1. Prorok PC, Miller AB. 1984. Screening for Cancer. I — General Principles on Evaluation of Screening for Cancer and Screening for Lung, Bladder and Oral, Cancer (UICC technical report series, vol. 78). Geneva: International Union Against Cancer.
2. Hakama M, Miller AB, Day NE. 1986. Screening for Cancer of the Uterine Cervix (IARC scientific publications no. 76). Lyon: International Agency for Research on Cancer.
3. Chamberlain J, Miller AB (eds.). 1988. Screening for Gastrointestinal Cancer Toronto: Hans Huber.
4. Day NE, Miller AB (eds.). 1988. Screening for Breast Cancer. Toronto: Hans Huber.
5. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.). 1991. Cancer Screening. Cambridge: Cambridge University Press.

1. The public health basis of cancer screening: principles and ethical aspects

Anthony B. Miller

1. Introduction

Cancer is a dread disease. It ably fulfills, at least in the mind of the public and of many physicians, two of the requirements for screening, namely, that the disease should be an important public health problem and that the consequences of untreatable cancer are dire [1]. However, we do not have a screening test for cancer; rather, we have a series of screening tests for different cancer sites. These tests use different approaches, possess varying sensitivity and specificity, and have produced varying evidence of efficacy and effectiveness. Thus screening for each cancer site has to be justified on its own merits and must therefore be evaluated using rigorous approaches, especially randomized controlled trials [2].

Screening for cancer is based on the assumption that early detection of cancers in the "detectable preclinical phase" (DPCP) [3] increases the chances of cure and thus will result in a reduction in mortality from the cancer in question. This assumption is followed naturally by the further assumption that the earlier in the DPCP the cancer (or its precursor) is found, the greater the reduction in mortality. Yet the empirical evidence supporting these assumptions is weak, and they can be challenged for a number of cancer sites [4].

The difficulty in evaluating screening is well recognized, especially the biases that relate to utilization of survival of screen-detected cases as a surrogate outcome for mortality (i.e., lead time, length bias, selection bias, and overdiagnosis bias) [5]. Some "lip service" is usually paid to these biases, but for those cancers for which clinical trials have demonstrated efficacy of screening in reducing mortality (breast and colon) (see chapters 5 and 6, this volume), death from the cancer is not abolished, and a high proportion of those who do not die of the cancer would not have died in the absence of screening. It is the small proportion who would die of their disease in the absence of screening but who, if their cancer is screen-detected, go on to die of another condition later who really benefit from screening. The fact that this component may be so small as to be undetectable was first demonstrated by the lung cancer screening trials, even though a number of small lung cancers were diagnosed earlier as a result of the screening tests used [2]. Further, there is

now evidence that at least some of the benefit from cervical cancer screening was derived before the initiation of cervical cytology screening by virtue of the earlier detection of cancers, rather than carcinoma in situ or "earlier" precursors [6].

2. Principles of screening

By definition, screening is offered to those who do not suspect that they may have a disease. This is subtly different from being asymptomatic. Symptoms may be revealed by careful questioning, related to the organ of interest, that may not be regarded by the screenee as being related to a possible disease. This approach became a subject of controversy in relation to the Canadian National Breast Screening Study (NBSS); although the frequency of disease was greater in those with breast symptoms, the majority of symptoms would not normally be related to breast cancer (e.g., premenstrual breast pain) [7,8].

For a public health program to be regarded as appropriate, it should normally be necessary to demonstrate first the efficacy of the approach and then its effectiveness as applied in the population. Both these components should normally be evaluated through randomized controlled trials, though less stringent methodologies may suffice for effectiveness once efficacy is demonstrated. The components of effectiveness include, in addition to efficacy, compliance both with the recommended screen and with the subsequent diagnostic maneuvers for those with an abnormality on the screen. It is not at all clear that the artificial circumstance of a controlled trial is always appropriate for measuring these important parameters. Further, randomized controlled trials are expensive and of long duration, and although they remain critical in determining the primary effect of screening in reducing mortality, new approaches to their design are being developed that will enable many secondary issues to be addressed more economically (see chapter 2 in this volume).

For both full effectiveness and efficiency of a public health intervention, an organized program is essential. The components of an organized screening program have been appreciated for over a decade [9], yet rarely in North America has it been possible to introduce such organization, especially for cervical screening [10]. These components are

- identifiable individual women in the target population;
- measures to guarantee high coverage and attendance, such as a personal letter of invitation;
- adequate field facilities for taking the smears and adequate laboratory facilities to examine them;
- an organized quality-control program for taking smears and interpreting them;
- adequate facilities for diagnosis and for appropriate treatment of confirmed neoplastic lesions;
- a carefully designed and agreed referral system for management of any

abnormalities found and for providing information about normal screening tests; and
- organized evaluation and monitoring of the total program [9].

Fortunately, for breast screening, at least in Europe and Canada, the necessity of organized programs has been recognized since the initiation of the population programs.

3. The ethics of screening

In medical practice, the special nature of the relationship between a patient and his or her physician has given rise to a core of ethical principles governing this relationship. Further, it has been recognized that special issues arise when a patient becomes the subject of research, which is superimposed on the patient's search for and receipt of appropriate medical care. It has not always been appreciated, however, that screening creates a new spectrum of issues that may require more restrictive boundaries of behavior than normally apply in medical practice. The crucial distinction between screening and normal medical diagnosis and care is that the provider of screening initiates the process, not the individual who is the subject of screening. This is true whether screening is initiated by governments, cancer societies, or public health units (sometimes described as "mass" or "public health" screening), or whether screening is carried out by the physician in his office (sometimes called "case-finding"). When a patient goes to see a physician for diagnosis of and hopefully relief from a symptom, or for treatment of an established condition, the physician is required to exercise his or her skills only to the extent that knowledge is currently available. In screening, however, those who are approached to participate are not patients, and most of them do not become patients. The screener believes that as a result of screening, the health of the community will be better. This does not necessarily imply that the condition of every individual screened will be better; indeed, in some circumstances, some individuals included in a screening program will be disadvantaged. Yet screening is often promoted as implying a benefit to every screenee. At the very least, therefore, those planning to introduce screening have an ethical responsibility to be able guarantee an overall benefit to the community. This has to be coupled with the responsibility to minimize by all possible means the harm that could accrue to some participants. These responsibilities imply that if valid evidence is not available from properly conducted research studies on the effectiveness of screening, screening programs should not be offered other than in the context of a properly designed experiment with validly constituted informed consent. This ethical imperative is perhaps particularly acute for those identified as gene carriers in high-risk families, a problem addressed by Foulkes and Narod (chapter 12, this volume).

Informed consent is a relatively recent innovation in randomized trials of screening. For breast screening, among those trials that have reported mortal-

ity results, only the NBSS required informed consent. However, for the newly planned or operational prostate screening trials, the majority require informed consent [11]. This has led to a recognition that randomized trials with informed consent evaluate the efficacy of screening; effectiveness trials dependent on randomized invitations in a defined population would be almost impossible with informed consent, so the measurement of effectiveness will tend to require other approaches in the future, such as quasi-experimental studies.

Those responsible for screening programs have the ethical responsibility to ensure that quality control of the screening tests is maintained, that the effectiveness of the programs is continually monitored [9], and that the specificity of the test is high [2]. The inability to guarantee overall benefit and lack of disadvantage to those screened led to the proscription of mammography in women under the age of 50 in the U.S. Breast Cancer Detection Demonstration Projects in the absence of certain specific indications for mammography [12]; and the continued lack of evidence of benefit has led most breast screening programs in Canada and outside the United States to be restricted to women age 50 or more.

There are some additional ethical issues related to the organization of screening. The first is to reduce unnecessary anxiety to the minimum. This is clearly a major responsibility in cancer screening programs requiring both a reduction in false-positive test results and the avoidance of overdiagnosis. The second is to ensure that if an abnormality is identified by screening, appropriate facilities are available for its diagnosis and treatment. Although this issue is a particular difficulty of some screening programs in developing countries, it has also arisen in programs in technically advanced countries [13]; (see also chapter 5, this volume). The third is to ensure that screenees with possible abnormalities do return for diagnosis and management. It has become apparent that failure to do so is one of the reasons for failure of screening programs for cancer of the cervix, even in developed countries [10,14].

A further ethical issue concerns the extent to which the offer of screening in a community could divert resources from other, more important, health care programs. This could be a particular problem for developing countries. There is an ethical responsibility to distribute limited resources equitably across the total community in order to obtain maximal benefit. Under certain circumstances, the offer of screening could diminish the overall level of health in a community, if it resulted in less resources being available for other diseases. However, Hakama (chapter 4, this volume) points out that a well-organized program, with efficient utilization of resources, could promote equity.

4. Advances in screening relevant to public health

In the 12 years of the UICC Project on screening that I chaired [5,9,15–17], we always deemed it appropriate to conclude our review of the "state of the art"

of screening by stating whether, in our view, screening for the relevant site was "applicable as public health policy." If our answer was "yes," we meant to imply that there was good evidence for the efficacy of screening and no particular expectation of ineffectiveness in the public health sense, providing organized programs as defined earlier could be introduced. As is apparent from the sites considered in this volume, the answer to our question is an unequicocal yes only for breast for women over the age of 50 and for cervix. Yet even for these sites, there are well-documented failures to reach the population at risk; in North America, even after many years of effort for cervix cancer screening, we may still be experiencing at least twice the irreducible minimum of disease [10]. For lung, neuroblastoma in children, and stomach cancer screening, the answer to the question appears to be no; the answer is probably still uncertain for colorectum, ovary, skin, and genetic screening. However, the reasons for the uncertainty differ. For colorectum, although Mandel (chapter 5, this volume) ponts out that there is good evidence for the efficacy of annual screening, he accepts the high cost and the problem with relative lack of specificity of the test. He also points out that biennial screening (being assessed in Europe) could halve some of these costs, yet his own trial in Minnesota failed to find benefit from biennial screening. Thus, there is still room for concern over effectiveness issues such as compliance, specificity, and cost for the fecal occult blood test. For ovary, one trial is in progress, yet it seems likely that issues related to low prevalence of the cancer and validity of the screening test will continue to direct public health policy decisions towards the negative. For melanoma, we have no trials, and although they are vigorously advocated by Elwood (chapter 10, this volume), none seems likely to be forthcoming in the immediate future; therefore, we have to allow those countries that have decided to screen to collect the observational data while the rest of us try to promote prevention (sun avoidance). For genetic screening (chapter 12, this volume) the issues are ethical concerns and the high cost of the tests, together with the likely inability of the approach to affect cancers in general.

Of particular importance, however, is the chapter by Bernstein and his colleagues (chapter 11, this volume) on the Quebec neuroblastoma experience. This is an excellent demonstration of one of the biases of screening, namely, overdiagnosis bias. The chapter is perhaps an opportune reminder that such a bias can affect other sites as well. We have good reasons to be concerned about it for preclinical abnormalities of the cervix, but it may also affect our evaluation of breast screening, as it almost certainly does for prostate and melanoma screening as well.

Further, for many if not all sites, Krahn and Naglie (chapter 3, this volume) remind us that there are many facets to a good economic analysis, and that so far, few of the published analyses include all the factors they recommend. Indeed, it seems probable that for many advocated screening tests recommended by some organizations as policy, a full economic analysis would indi-

cate that they are not beneficial. The evidence available on breast screening in women under the age of 50 (Moss, chapter 6, this volume) suggests that, in this instance, screening in the public health sense cannot be justified [18].

5. Conclusion

Although screening is an important component of cancer control, there are many barriers to it achieving its full potential in the public health sense. There are also many ethical issues to consider, in relation both to research on screening, and to its application. There seems little reason to change the (pessimistic to some) estimate of the Year 2000 Committee of a decade ago, which concluded that screening could only contribute about 3% to anticipated reduction in cancer mortality by the year 2000 [19]. All this benefit would accrue to women from breast screening for women age 50–69 and cervix screening for women age 20–69. There is no reason at present to believe that this was an underestimate.

References

1. Miller AB. 1982. Fundamental issues in screening. In Schottenfeld D, Fraumeni JF Jr (eds.), Cancer Epidemiology and Prevention. Philadelphia: W.B. Saunders, pp. 1064–1074.
2. Prorok PC, Chamberlain J, Day NE, Hakama M, Miller AB. 1984. UICC Workshop on the evaluation of screening programmes for cancer. Int J Cancer 34:1–4.
3 Cole P, Morrison AS. 1980. Basic issues in population screening for cancer. J Natl Cancer Inst 64:1263–1272.
4. Miller AB. 1994. Screening for cancer: Is it time for a paradigm shift? Ann R Coll Physicians Surgeons Canada 27:353–355.
5. Miller AB. 1985. Principles of screening and of the evaluation of screening programs. In Miller AB (ed.), Screening for cancer. Orlando, FL: Academic Press, pp. 3–24.
6. Ponten J, Adami HO, Bergstrom R, Dillner J, Friberg LG, Gustafsson L, Miller AB, Parkin DM, Sparen P, Trichopoulos D. 1995. Strategies for global control of cervical cancer. Int J Cancer 60:1–26.
7 Miller AB, Baines CJ, To T, Wall C, et al. 1992. Canadian national breast screening study: 1. Breast cancer detection and death rates among women age 40–49 years. Can Med Assoc J 147:1459–1476.
8. Miller AB, Baines CJ, To T, Wall C, et al. 1992. Canadian national breast screening study: 2. Breast cancer detection and death rates among women age 50–59 years. Can Med Assoc J 147:1477–1488.
9. Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC. 1985. Evaluation of screening programmes for gynaecological cancer. Br J Cancer 52:669–673.
10. Miller AB. 1995. Editorial: Failures of cervical cancer screening. Am J Public Health 85:761–762.
11. Auvinen A, Rietbergen JBW, Denis LJ, Prorok PC, Schröder FH, for the International Prostate Cancer Screening Trial Evaluation Group. In preparation. Prospective evaluation plan for randomized trials of prosate cancer screening.
12. Beahrs OH, Shapiro S, Smart S. 1979. Report of the working group to review the National Cancer Institute–American Cancer Society Breast Cancer Detection Demonstration Projects. J Natl Cancer Inst 62:641–709.

13. Mandel JS, Bond J, Snover D, Williams S, Bradley M, Walker C, Schuman LM, Gilbertsen V. 1988. The University of Minnesota's Colon Cancer Control Study. In Chamberlain J, Miller AB (eds.), Screening for Gastrointestinal Cancer. Toronto: Hans Huber, pp. 17–24.
14. Chamberlain J. 1986. Reasons that some screening programmes fail to control cervical cancer. In Hakama M, Miller AB, Day NE (eds.), Screening for Cancer of the Cervix Uteri. IARC Scientific Publications No. 76. Lyon: International Agency for Research on Cancer, pp. 161–168.
15. Chamberlain J, Day NE, Hakama M, Miller A, Prorok PC. 1986. UICC workshop of the project on evaluation of screening programmes for gastrointestinal cancer. Int J Cancer 37:329–334.
16. Day NE, Baines CJ, Chamberlain J, Hakama M, Miller AB, Prorok P. 1986. UICC project on screening for cancer: report of the workshop on screening for breast cancer. Int J Cancer 38:303–308.
17. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC. 1990. Report on a workshop of the UICC Project on evaluation of screening for cancer. Int J Cancer 46:761–769.
18. Miller AB. 1993. Canadian National Breast Screening Study: Public health implications. Can J Public Health 84:14–16.
19. Greenwald P, Sondik EJ (eds.). 1986. Cancer control objectives for the nation: 1985–2000. NCI Monographs, No. 2.

2. The theoretical basis for cancer screening

Nicholas E. Day

1. Introduction

Screening programs for cancer demand a major allocation of public health resources. The evidence required to justify the introduction of such programs therefore needs to be unassailable, based on randomized trials with mortality reduction from the cancer in question as the primary endpoint. These trials, which could be called *primary*, are of necessity large and of long duration, since the yearly risk of dying from a cancer is small in absolute terms for almost all cancers at almost all ages. For example, there are few female populations for which the yearly risk of dying from breast cancer is more than 0.1% at any age.

The cost of these trials raises two issues. Firstly, the information needed to justify initiating a primary trial must be determined. Secondly, once the effectiveness of a screening modality has been demonstrated and population screening introduced, a range of proposed modifications will need to be assessed that represent the fine tuning needed to optimize benefits. It would be perverse to undertake trials to resolve subsidiary issues (which one could call *secondary* trials) that are larger than primary trials. For both these issues, one needs a theoretical understanding of the process of screening in order to predict expected mortality benefits from earlier results.

Once a screening modality is introduced as a public health measure, further issues arise, many related to ensuring that the mortality reduction anticipated on the basis of the primary trials will be delivered by the population screening program. For this purpose, one needs to define a set of measures for monitoring the program, based on the early results as they emerge, that is predictive of the long-term mortality outcome [1]. Establishing the adequacy of such monitoring requires, as before, a theoretical understanding of the screening process. In this description of the screening process, three issues will be of concern:

- what information is required to initiate primary trials;
- how secondary trials should be designed; and
- how measures can be developed for the adequate monitoring of population screening programs.

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

Several approaches in the past have developed so-called "deep" models of cancer screening [2,3]. In this chapter, a different approach is adopted that is more closely based on observable epidemiological quantities, in which the logical relationships between these quantities are emphasized.

2. The screening process: screening for early invasive cancers

2.1. Initial considerations

In screening for cancer, there are two distinct targets depending on the natural history, as currently understood, of the malignancy. For some cancers, notably of the cervix, and for sigmoidoscopy screening for large bowel cancer, a preinvasive condition is recognized, of long duration and through which most invasive cancers are thought to pass. This preinvasive condition is the primary target of screening. For cancers at other sites, including the breast, ovary, colorectum (fecal occult blood screening), prostate, and stomach, the primary target of screening is early invasive cancer. Preinvasive lesions may be detected, such as ductal carcinoma in situ of the breast, but their role in mortality reduction is at most marginal. This latter situation is considered first, and attention will focus on breast screening.

When the screening test is applied, preclinical invasive lesions are identified. Two questions arise:

- When would these lesions have surfaced clinically? Addressing this question introduces the concepts of sensitivity, specificity, lead time, and sojourn time, which are essentially descriptive measures.
- What change in prognosis has been achieved by advancing the time of

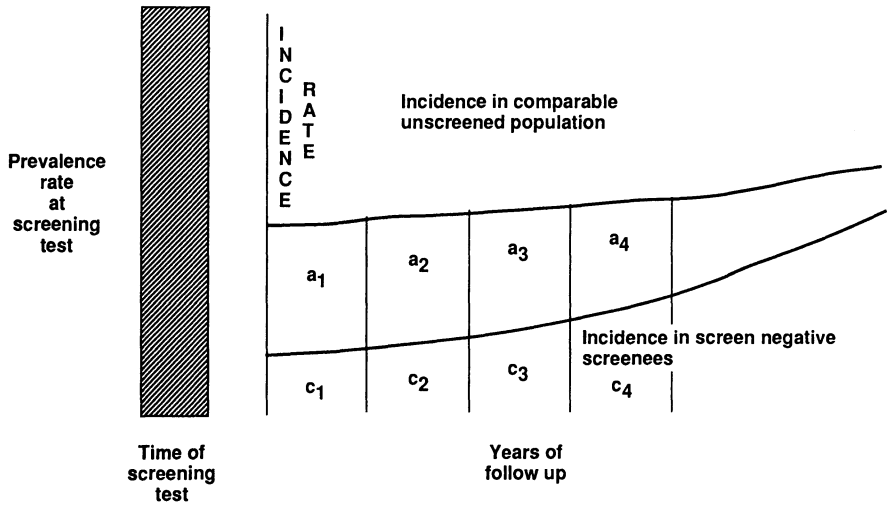


Figure 1. The cancers diagnosed at and after a single screening test.

diagnosis? This question addresses the fundamental issue of the disease's natural history, and for its resolution requires the definition of predictive measures.

In figure 1, the results of a single screening test and subsequent follow-up are displayed. Among those screened, a proportion *P* have a preclinical invasive cancer detected. Among those classed as negative, the incidence rate of clinical cancers rises with elapsed time since the screening test until it approaches the rate that would have been observed if the population had not been screened. This rate is immediately derived from the control group in randomized trials. In nonrandomized studies, assumptions may have to be made on the degree of self-selection for screening. No assumption is made that this rate is constant.

In figure 1, area c_i represents the clinical cases arising among those screened negative in the interval (i-1, i) years after screening. Area $a_i + c_i$ represents the clinical cancers that would have arisen in this interval if no screening had occurred. Area a_i thus represents the cases that would have arisen in this interval in the absence of screening, but were detected at the screening test. One expects c_i to increase with i, this increase representing the increasing number of cancers that were undetectable at the screening test.

The traditional, elementary table that defines sensitivity and specificity takes no account of time:

		Gold standard	
		+	−
Test result	+	a	b
	−	c	d

$$\text{Sensitivity} = \frac{a}{a+c} \qquad \text{Specificity} = \frac{d}{b+d}$$

$$\text{Predictive value positive} = \frac{a}{a+b}$$

In these terms, if test-positive cases are taken as all those detected at the prevalence screen, it is impossible to define an equivalent group of gold standard positive cases, since one does not know when they might have occurred.

A different approach is required taking account of time. Definitions given earlier introduce an instantaneous sensitivity, together with the distribution of sojourn times [4,5]. More simply, and of greater operational utility, one can define the *year i sensitivity* (Sn_i) as

$$Sn_i = a_i / (a_i + c_i)$$

It represents, among the cancers that in the absence of screening would have surfaced clinically in the ith year after screening, the proportion that were detected on screening.

A series of values Sn_1, Sn_2, Sn_3, . . . are thus defined, the year 1 sensitivity, the year 2 sensitivity, etc., representing the ability of the screening test to detect cancers that would surface increasingly further in time after the test. These values are also described as proportionate incidence rates of interval cancers, but this term ignores the essential meaning of these ratios in terms of sensitivity.

This approach can now be extended to define specificity and positive predictive value. A horizon needs to be chosen — n years, say. The basic 2×2 table becomes

		Clinical cancers surfacing in n years		
		+	−	
Screening test	+	$\sum_i a_i$	b	P
	−	$\sum_i c_i$	d	$N - P$
		$\sum_i (a_i + c_i)$	$b + d$	N

where the summation is from 1 to n, P is the number of screen-detected cancers, and N is the total population screened.

The n year positive predictive value is then

$$\sum_{i=1}^{n} a_i / P$$

The complement of this quantity gives the proportion of screen-detected cancers that would not surface clinically within n years.

The n-year false-positive rate (1–specificity) is given by

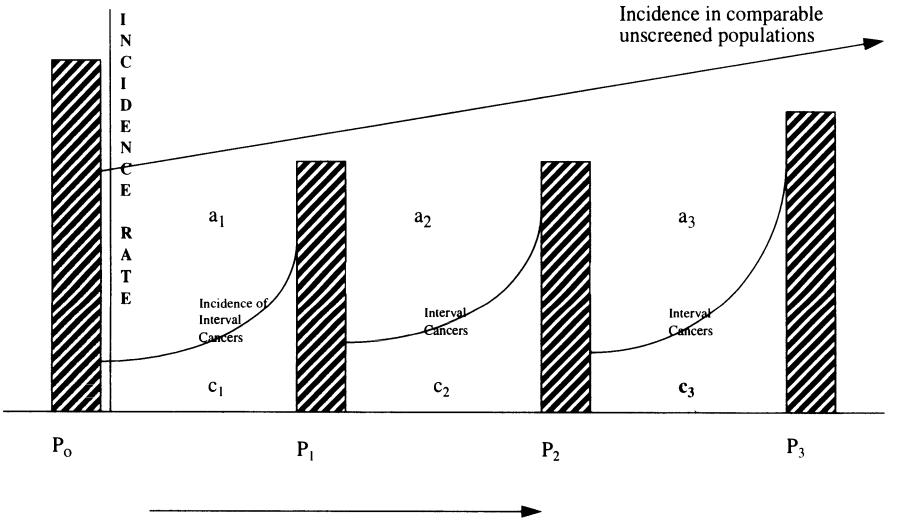
$$b/(b+d)$$

The amount of lead time gained by screening for cancers that would have surfaced clinically can also be obtained directly from figure 1. In the first year postscreening, a_1 cancers would have gained approximately 1/2 year of lead time, a_2 cancers a corresponding $1^1/_2$ years, and so on. In this way, the distribution of lead time associated with a single screening test can be built up. Knowledge of the lead time distribution is of little utility, however, if one cannot estimate the change in prognosis that is induced. To assess the change in prognosis, and hence the reduction in mortality, that is achievable through screening, two equivalent sets of cancers need to be compared, one arising clinically in the absence of screening, the second comprising cancers diagnosed when a screening program is in operation. With a single application of a screening test, difficulties arise. Choosing a time interval of n years, from figure 1, the two groups of cancers consist of $\Sigma (a_i + c_i)$, the cancers occurring without screening, and $P + \Sigma c_i$, the cancers occurring with a single screening (both summations from 1 to n). For these two sets to be equivalent, one needs Σa_i to

be equivalent to P, i.e., the incidence gap must be equal to the number of screen-detected cancers minus the positive predictive value equal to unity. For this to occur, n may have to be large, in which case both sets will be dominated by the term Σc_i, the interval cancers, the prognosis of which is unchanged by screening. If a smaller value is chosen, then the positive predictive value may be substantially less than unity, implying an excess of cancers in the set occurring with a single screening test. In either case, the effect on prognosis will be underestimated. This dilemma is avoided if periodic screening is considered.

2.2. Periodic screening: Program sensitivity and the unbiased set

In figure 2, the results are represented schematically for the cancers occurring when a population undergoes periodic screening with a fixed interscreening interval of T years [6]. Among individuals presenting regularly for screening, two types of cancer occur, namely, screen-detected and interval cancers. Clearly, the relative proportion of the two types and their prognostic characteristics will be the determinants of the effectiveness of screening with that interval. The rate of interval cancers, as a proportion of what would have occured without screening, will be the sum of the yearly sensitivities as defined



P_o Cancers detected at the prevalence screen

P_i Cancers detected at the 'i'th incident screen

c_i Interval cancers diagnosed in the 'i'th inter screening interval

a_i The incidence gap, in the 'i'th interscreening interval

Figure 2. The cancers occurring in a population undergoing periodic screening, at an interscreening interval of T years.

early. It is then convenient to define a pragmatic measure of sensitivity, the *program sensitivity*:

$$Programe\ sensitivity = \left(\sum_{1}^{T} Sn_i\right) \Big/ T$$

i.e., the average yearly sensitivity over the screening interval of length years. Over a single screening cycle, this sensitivity will be given by *Program Sensitivity* = $a_i/(a_i + c_i)$, provided that the incidence in the unscreened population is close to constant over the interval — a reasonable assumption for intervals of up to three years.

If we designate as a screening cycle the time interval from immediately after one screening test to immediately after the next [7], then the cancers diagnosed during a screening cycle constitute an unbiased set [8,9]. This set is unbiased in the sense that it is almost unaffected by length bias and that the cumulative incidence over the interval, including the screening test at the end, is equal to the cumulative incidence over the same interval in an equivalent unscreened population. The result derives from the fact that the unbiased set consists of the newly incident cancers in this time period, but with a diagnostic threshold defined by screening rather than by clinical surfacing. Figure 3

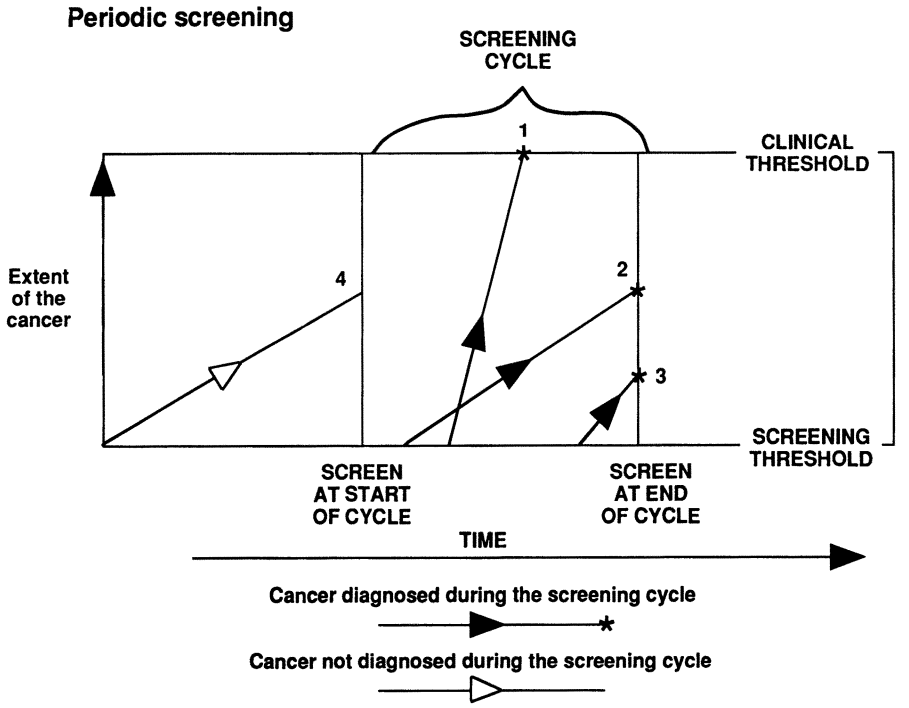


Figure 3. Schematic representation of cancers diagnosed during a screening cycle, demonstrating how they form an unbiased set.

provides a graphical illustration. Algebraically, this lack of bias has been demonstrated under specific assumptions, including high sensitivity and specificity and no regression [10].

The last requirement is critical, namely, that invasive cancers do not regress. Under most circumstances, spontaneous early regression of an invasive cancer would be considered possible but very rare. In premenopausal breast cancer, with screening cycles that span the menopause, appreciable regression of small invasive lesions seems to occur [11].

The concept of the unbiased set over a screening cycle provides the fundamental insight into how screening works. For most types of cancer, tumor characteristics have been identified that determine survival; tumor stage is usually a bald summary of some of these measures, based on size of the primary tumors and extent of disease. Other characteristics, based on histopathological, cellular, or molecular markers, may give additional prognostic information [8]. Comparison of the distribution of these characteristics between the set of cancers diagnosed over a screening cycle and those diagnosed in an unscreened population gives a direct estimation of the effect of screening in improving prognosis [12]. Comparison between the cancers detected during screening cycles of different lengths, with different relative proportions of interval and screen-detected cancers, will provide an estimate of the effect on prognosis of different screening intervals [7].

Table 1 demonstrates from the Swedish two-county study the results over the first three screening cycles for women aged 40–69 at entry, comparing with respect to tumor size, nodal status, and malignancy grade the study group with

Table 1. Swedish two-county study (age group 40–69 on entry to the study)

	Study group	Control group
Distribution of size at diagnosis among incident cancers		
N =	704	590
Size (mm):		
1–9	18%	7%
10–14	22%	15%
15–19	21%	20%
20–29	23%	29%
30–49	11%	20%
50+	5%	9%
Nodal status among incident tumors		
N =	670	558
Negative	68%	54%
Positive	32%	46%
Malignancy grade distribution among incident tumors		
N =	600	493
Grade:		
I	22%	16%
II	38%	36%
III	40%	48%

the control group (those refusing screening in the study group are included, since they have counterparts in the control groups; the prevalence screening round is excluded).

For prediction of effectiveness in reducing mortality, the change in prognostic variables needs to be converted into change in survival. For this purpose, data are needed on survival for each category of the prognostic variables. Combining the survival curves with the distribution of time of diagnosis for each prognostic category will provide predicted mortality into the future, from which bias due to lead time and length bias has been removed [7,12]. Two difficulties arise: (1) survival by category of the prognostic variable may be different between screen-detected and clinically diagnosed cancers, and (2) the most favorable prognostic categories seen in a screen-detected series may be virtually unrepresented in a clinical series. To resolve both problems, long-term survival data are needed from both screening and clinical series. The implication is that primary screening trials need to be completed before predicted mortality can be used as a reliable means of evaluation.

In the Swedish two-county study, using the results in table 1 and the observed survival data, a comparison can be made between the predicted effects of screening and the observed data. These are given in table 2 and show good agreement [12]. It should be noted that the confidence intervals for the predicted mortality reduction are narrower.

The unbiased set consists of interval and screen-detected cancers. Survival of the former, when it has been reported, is often similar to that of cancers in an unscreened group. Improved survival in the unbiased set is therefore determined mainly by the improved survival of the screen-detected lesions. When screening detects predominantly preinvasive lesions, this survival should be almost that of the healthy population, and then the effectiveness of screening depends entirely on the rate of interval cancers.

Throughout the above discussion, the initial prevalence screen has been ignored. It appears in no screening cycle, as the cycle has been defined. The reason is that the length bias seen at the first screening test is too great for it to be removed by the procedure described above. For the two cancers for which extensive information is available, namely, breast and cervix, many

Table 2. The screening-associated reduction in breast cancer mortality observed and predicted on the basis of tumor size, nodal status, and malignancy grade, by age group [12]

	Reduction in breast cancer mortality	
Age group	Observed RR (95% CI)[a]	Predicted RR (95% CI)[a]
40–49	0.87 (0.54–1.41)	0.96 (0.73–1.29)
50–74	0.66 (0.54–0.81)	0.71 (0.63–0.80)

[a] Relative risk (95% confidence interval).

prevalence screen-detected cancers have sojourn and lead times considerably longer than the accepted interscreening interval. The accumulation of lesions with a long sojourn time seen at the initial screening test has no counterpart at later screening tests. For breast cancer, this essential length bias at the prevalence screen is reflected in the inability of the usual prognostic factors to explain fully the good survival of prevalence screen cancers [8].

3. Model-based evaluation

The application of the concepts outlined above concerning the three issues raised in the introduction is straightforward.

3.1. The development of primary trials

For societies with a rational approach to the allocation of research resources, given the cost and size of primary screening trials, objective evidence that the trials have a reasonable chance of demonstrating a substantial mortality benefit of public health interest should be a requirement before a full trial is launched. When screening detects early invasive lesions, this requirement is best met by the provision of data on a complete screening cycle — in particular, the improvement in the distribution of prognostic factors seen over a screening cycle compared to the distribution seen in an unscreened control group. In other words, what does screening detect and what does it miss? The exception is when screening identifies preinvasive lesions. In this situation, the crucial quantity is the rate of interval cancers, for which approximate estimates are required. The development of the trial of flexible sigmoidoscopy [13] illustrates the point.

Typically, the sample size required to provide adequate estimates over one screening cycle of the prognostic factor distribution is 20% to 30% of that required for a full primary trial. Demonstrations of a reasonable probability of a successful outcome should be a pre-requisite for the funding of primary trials.

3.2. The design of secondary trials

In many circumstances, secondary trials would be considered only after successful primary trials have been completed. Information would then be available on the survival associated with different categories of the prognostic variables for the screen-detected and clinically diagnosed cancers. The potential to use predicted mortality as an endpoint requires the identification of adequate prognostic factors to be used as surrogates for mortality from the cancer of interest. A helpful definition of an adequate surrogate has been given by Prentice [14] and extended by Freedman et al. [15]. This states that if λ is the mortality rate, x is the treatment or screening modalities being com-

pared, and t is time since diagnosis, then a set S of variables constitutes adequate surrogates for mortality if

$$\lambda(t|S,x) = \lambda(t|S)$$

i.e., the effect of treatment on mortality is entirely explained by the effect of treatment on the surrogates S. These would usually be tumor characteristics determined at the time of diagnosis.

If, on the basis of a primary trial, one is satisfied that an adequate set of surrogate variables can be defined from which predicted mortality could be used as a trial endpoint, then the design of secondary trials can be based on the characterization of the cancers seen over completed screening cycles on the different arms of the trial. The benefits can be outlined as follows:

Suppose the set S of surrogate variables takes values 1, ... n, and the distribution of these values under tow treatment arms is

$$P(S = i|x = j) = q_{ij}$$

$i = 1, \ldots n$ and $j = 0, 1$.

Suppose also that the probability of death from the cancer in question for an individual in category i of S is P_i, independent of the treatment arm.

The parameter that the trial is designed to estimate is the hazard ratio between the two arms (hazard of death from the cancer of interest). The variance of the logarithm of the hazard ratio if observed mortality is the endpoint is given by

$$V_{\text{obs}} = \frac{1}{N}\left(\frac{1}{\sum P_i q_{i1}} + \frac{1}{\sum P_i q_{i0}}\right)$$

and, if predicted mortality is used,

$$V_{\text{pred}} = \frac{1}{N}\left(\frac{\sum P_i^2 q_{i1}}{\left(\sum P_i q_{i1}\right)^2} + \frac{\sum P_i^2 q_{i0}}{\left(\sum P_i q_{i0}\right)^2}\right)$$

V_{pred} cannot be greater than V_{obs} and typically might be three to five times smaller. An example is given in table 3 [7]. In other words, a trial could be three to five times smaller and provide the same precision. The trial would also, of course, be several times shorter, since the information required is available at diagnosis. This approach is being used in the U.K. to assess the relative benefit of different frequencies of screening for breast cancer. It is clearly an approach that might have wide application. An example is the proposed trial of breast screening for women under 50 (EUROTRIAL 40), which has breast cancer mortality as the primary endpoint [16]. As designed, the trial is seriously wasteful of resources and likely to be made redundant years before achieving a result. It has, however, the built-in capacity to answer a wider range of questions through the use of surrogate measures than can be

Table 3. Size distribution of all tumors diagnosed (screen-detected and interval cancers) during screening cycles of three-year and one-year duration, with size-specific 10-year death rates from the Swedish two-county study

	% of tumors for screening cycle		
Size (mm)	Three-year	One-year	10-year death rate
1–9	20	43	0.02
10–14	25	21	0.05
15–19	21	14	0.15
20–29	20	13	0.35
30–49	9	7	0.50
50+	5	2	0.70
% dying in 10 years	20	13	

Note: $V_{obs}/V_{pred} = 2.57$.

addressed with mortality as the endpoint and to obtain the answers much more rapidly.

It should be emphasized that the value of surrogate endpoints in the design of secondary trials is based on assumptions that need to be validated on each occasion, and this validation comes mainly from the results of previous trials.

3.3. Monitoring public health programs

When a screening program such as the breast screening program is introduced, the anticipated mortality reduction will not emerge for a number of years. To ensure that the program is on track to deliver the predicted benefit, or to identify shortcomings at an early stage so that remedial action can be taken, informative methods of monitoring the program are required. The measures chosen for monitoring purposes can be descriptive or predictive. Descriptive monitoring measures have been described for the U.K. Breast Screening Program since its inception [1,17,18]. These are based on straightforward considerations of the process of the program, as described in table 4. The rationale for their target was to mimic the result of the Swedish two-county trial, on which the British program was largely based. It was considered plausible that if the targets on these early measures were achieved, then the ensuing reduction in breast cancer mortality would compare with that of the two-county trial. No attempt was made to quantify the effect of not achieving these targets. Recently, an attempt has been made to make these measures more predictive [19] on the basis of interval cancer rates, and concurrently the NHSBSP targets have been made more stringent. A more complete approach to predictive monitoring of the U.K. program is being developed, however, based on the results of a completed screening cycle and the concept of the "unbiased set" (McCann, personal communication). This approach requires the results of the second round of screening, in order to define the results over an entire screening cycle (i.e., interval cancers from just after the prevalence

Table 4. Monitoring measures defined for the U.K. Breast Screening Program

	1989 [17]	1993 [18]
Compliance	70%	70%
Prevalence round referral rate	<10%	<7%
Prevalence round cancer detection rate	>5 per 1000	>5 per 1000
Prevalence round detection of cancer <10mm	>1.5 per 1000	>1.5 per 1000
Malignant: benign ratio	>1:3	>1:1
Interval cancers within 12 months	<6 per 10,000	<3 per 10,000
Rescreening cancer detection rate	—	>3.5 per 1000

Table 5. Predictive monitoring of a breast screening program using tumor size of the tumor characteristics (invasive cancers occurring in the first three-year screening cycle per 10,000 women invited, assuming a background rate of 20/10,000 women per year)

Tumor size (mm)	Interval cancers			Cancer detection rates at rescreening	Cancers in refusers	Total over a screening cycle
	Year 1	Year 2	Year 3			
1–9						
10–14						
15–19						
20–29						
30–49						
50+						
Total rate	4.8	8.8	12.8	21.6	12	60

screen to just before the second screening round) plus the cancers detected at the second screening round. Cancers diagnosed over this screening cycle in those refusing screening at the prevalence round are also included. These three sets of cancers together form the set of cancers on which predicted mortality is to be based, which is then compared with predicted mortality over an equivalent time period in the absence of screening.

Table 5 presents the details using artefactual results. It should be noted that the results of the prevalence round are not included, for reasons described earlier. The average survival in this table can then be compared with the average survival in an unscreened population. In table 5, for ease of presentation, tumor size is taken as the sole surrogate variable. The table presents simulated results for a population of 10,000 invited for screening, of whom 80% attend the prevalence round, leaving 20% as refusers. Interval cancer rates are taken from the U.K. program [19]. For ease of presentation, it is assumed that all women attending the prevalence round attend the first rescreening round. Using survival information by size from other sources, or from historical data in the same population, the size-specific incidence

rates over a screening cycle can be converted into predicted mortality, to be contrasted with the predicted mortality of a comparable unscreened population.

We should note from table 5 that the predicted reduction in mortality depends on three aspects of the screening results, namely, the compliance rate, the rate of interval cancers, and the size distribution among screen-detected cancers. With respect to the last, failure to detect small cancers at screening will not only increase the rate of interval cancers but also reduce the improved survival among those with screen-detected cancers. Most of the benefit of screening is derived from detecting cancers when small; if this number is reduced, the benefit is directly diminished.

4. Screening for preinvasive lesions

Similar concepts apply in this situation, with the simplification that screen-detected lesions will be (largely) of very good prognosis, but with the added complexity that the natural biology of the screen-detected preinvasive lesions may be poorly understood, and many may have little potential to progress. Many in fact may regress. The definition of specificity and positive predictive value given earlier is directly relevant to the estimation of the proportion of preinvasive lesions that may regress. The most comprehensive attempt to estimate regression rates in cervical cancer screening is based on the British Columbia cervical screening program [20]. One approach taken was to compute, as in figure 1, the number of invasive cancers that appear after screening compared to the number expected in the absence of screening. The difference indicates the number of screen-detected lesions that would have progressed to invasion, a number that can be compared with the total number of lesions that were screen-detected. This ratio is exactly the positive predictive value, as defined in section 2.1.

If one assumes that the screen-detected lesions, at least after the initial prevalence screen, contribute a negligible amount to the predicted mortality based on the lesion occurring in a screening cycle, then computation of predicted mortality depends only on the rate of interval cancers. Thus, in the justification of the development of large primary trials, emphasis should be placed on the evidence for interval cancer rates. One sees exactly this approach in the development in the U.K. of the trail of flexible sigmoidoscopy as a mass screening modality for the prevention of colorectal causes [13].

In cervical screening, the question of the design of primary or secondary trials of cytological screening is theoretical, since none has been or is likely to be undertaken. The question of trials to assess the value of Human Papilloma Virus (HPV) testing as an adjunct to cytology, in particular to improve specificity, will not be considered here. The evaluation of cytological screening has been based on observational data, with between-population comparisons and time-trend analysis providing evidence for the effectiveness of screening (a

Table 6. Screening for cancer of the cervix

a. Relative risk of invasive cervical cancer by year of the negative cervical smear, among women with at least two negative smears

Time since negative smear (year)	Relative risk
0–	0.07
1–	0.08
2–	0.12
3–	0.19
4–	0.36
5–	0.28
6–	0.63
10+	1.20
Never smeared	1.00

b. Reduction in cumulative incidence of invasive cancer over the age range 35–64, with different screening frequencies

Screening frequency (year)	%Reduction in cumulative rate
1	93.3
2	93.3
3	91.4
5	83.9
10	64.2

poor substitute for primary trials). The main secondary issue has been the evaluation of the relative effectiveness of different screening intervals. One approach, shown in table 6, has been by direct observation in case–control and cohort studies [21]. An alternative approach is through modeling the natural history [22–24]. Consistency of the two approaches has recently been demonstrated, as shown in figure 4 [25], but with the interesting additional point that estimates of the relative effectiveness of different screening intervals depend on whether occurrence of invasive disease or mortality from the disease is used as the endpoint. In the IARC [21] study, invasive disease was taken as the endpoint. Over a screening cycle (figure 2), the relationship between the length of the cycle and the average yearly incidence of invasive disease is given correctly by table 6.

To evaluate predicted mortality, however, one needs to proceed to the computation indicated in table 5, in which the size (or stage, in the case of cervical cancer) distribution is taken into account, and the stage-specific survival incorporated. Since most screen-detected lesions have an excellent prognosis, the main contribution to mortality from cervical cancer derives from the interval cancers, clinically invasive, that arise in a screening cycle. The reduction in mortality is then greater than the reduction in incidence, and the quantities in table 6 underestimate the reduction in mortality associated with different screening intervals.

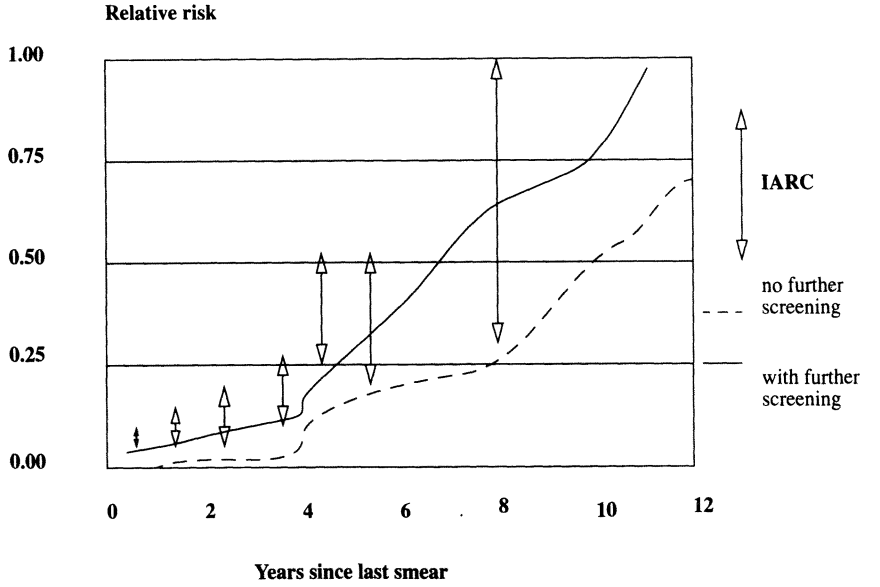


Figure 4. Relative risk of diagnosis of an invasive cervical cancer after the last of at least two consecutive negative cytological smears. Comparison of empirical results of the IARC study [21] and theoretical modeling [3].

5. Concluding remarks

Simple, observation-based models are required to extrapolate from the results of primary trials to the implementation of screening on a population basis. These models are necessary in order to understand how screening achieves a reduction in cancer mortality. This understanding can then form the basis for the evaluation of different implementation policies, for the design of feasible, relatively small and short-term secondary trials in which specific uncertainties are resolved, and for the development of effective program monitoring strategies that have predictive power. This chapter has attempted to demonstrate how development and application of these models for breast and cervical screening can lead to a coherent approach to the development of a public health strategy, including the design of research and development activities that should be an essential component of such a strategy. The chapter has not considered the problems that arise at an earlier stage of development of a screening modality. For ovary cancer screening, for example, the reluctance to initiate large primary trials at present is founded on more basic considerations of harm and benefit than are considered here.

References

1. Day NE, Williams DRR, Khaw KT. 1989. Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer 59:954–958.

2. Eddy DM. 1980. Screening for cancer: Theory, analysis and design. Englewood Cliffs, NJ: Prentice-Hall.
3. van Oortmarssen GJ, Habbema JDF. 1995. Duration of preclinical cervical cancer and reduction in incidences of invasive cancers following negative Pap smears. Int J Epidemiol 24:300–307.
4. Day NE. 1985. Estimating the sensitivity of a screening test. J Epidemiol Commun Health 39:364–366.
5. Paci E, Duffy SW. 1991. Modelling the analysis of breast cancer screening programmes: Sensitivity, lead time and predictive value in the Florence district programme (1975–1986). Int J Cancer 20:852–858.
6. Day NE, Walter SD. 1984. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics 40:1–14.
7. Day NE, Duffy SW. 1996. Trial design based on surrogate endpoints — application to comparison of different screening frequencies. J R Stat Soc Series A 159:49–60.
8. Duffy SW, Tabar L, Fagerberg G, et al. 1991. Breast screening, prognostic factors and survival results from the Swedish Two County study. Br J Cancer 64:1133–1138.
9. Walter SD, Day NE. 1983. Estimation of the duration of a pre-clinical disease state using screening data. Am J Epidemiol 118:865–885.
10. Walter SD, Stilt LW. 1987. Evaluating the survival of cancer cases detected by screening. Stat Med 6:885–900.
11. Tabar L, Fagerberg G, Chen HH, et al. 1995. Screening for breast cancer in women aged under 50: mode of detection, incidence, fatality and histology. J Med Screening 2:94–98.
12. Tabar L, Fagerberg G, Chen HH, et al. 1995. Efficiency of breast screening by age. Cancer 75:2507–2517.
13. Atkin WS, Cusick J, Corthoves JMA, et al. 1993. Prevention of colorectal cancer by once only sigmoidoscopy. Lancet 341:736–740.
14. Prentice RL. 1989. Surrogate endpoints in clinical trials: definition and operating criteria. Stat Med 8:431–440.
15. Freedman LS, Granbard BI, Schatzkin A. 1992. Statistical validation of intermediate endpoints for chronic diseases. Stat Med 11:167–178.
16. Eckhardt S, Badellino F, Murphy GP. 1994. UICC meeting on breast screening in premenopausal women in developed countries. Br J Cancer 56:1–5.
17. NHSBSP. 1989. Quality assurance guidelines for mammography. U.K. Department of Health.
18. NHSBSP. 1993. Objectives for the breast screening programme. U.K. Department of Health.
19. Day NE, McCann J, Camilleri-Ferrante C, et al. 1995. Monitoring interval cancers in breast screening programmes: the East Anglian experience. J Med Screening 2:180–185.
20. Boyes DA, Morrison B, Knox EG, et al. 1982. A cohort study for cervical cancer screening in British Columbia. Clin Invest Med 5:1–29.
21. IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. 1986. Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policy. Br Med J 113:214–226.
22. Gustafsson L, Adami HO. 1989. Natural history of cervical neoplasia: consistent results obtained by an identification technique. Br J Cancer 60:132–141.
23. Gustafsson L, Adami HO. 1990. Cytologic screening for cancer of the uterine cervix in Sweden evaluated by identification and simulation. Br J Cancer 61:903–908.
24. Gustafsson L, Adami HO. 1990. Optimization of cervical cancer screening. Cancer Causes Control 3:125–136.
25. Oortmarssen GJ van, Habbema JDF, van Maas PJ, et al. 1990. A model for breast screening. Cancer 66:1601–1612.

3. Principles of economic evaluation in cancer screening

Murray Krahn and Gary Naglie

1. Introduction

Should high-risk women under the age of 50 be screened for breast cancer? What about women over 70? How often should cervical cytology be assessed? Should asymptomatic men have a prostate-specific antigen performed? Who, if anyone, should receive CA125 and transvaginal ultrasound screening for ovarian cancer? How often should stool occult blood testing and/or sigmoidoscopy be performed to screen for colorectal cancer?

These are vexing questions not only for clinicians but also for those who organize and pay for health care. Formulating rational policies for cancer screening is an extraordinarily complex and difficult undertaking because the diseases are complicated, screening techology evolves quickly, and the data about screening efficacy are often incomplete. In addition, in this era of constrained health resources, we are obliged to think about cost as well as clinical benefit. Even wealthy nations, squeezed between the Scylla of exponential technology growth and the Charybdis of shrinking health care resources, must make hard choices about which technologies they can afford.

Economic evaluation, or "cost-effectiveness analysis," is a methodology that is increasingly being used to help clinicians and policy makers think about these hard choices. Economic evaluation is the "comparative analysis of alternative courses of action in terms of both their costs and consequences" [1]. In measuring the effects of health programs on both resource consumption and health, economic evaluation allows us to see how much health care "bang" we're getting for our "buck." By expressing the costs and consequences of programs in standardized units (e.g., dollars, quality-adjusted life years), it provides us with an intellectual framework that allows comparisons of competing programs not only within but also across clinical domains.

In this chapter, we discuss the principles of economic evaluation in the context of screening for cancer. Our intent is to provide both a reader's guide to the literature and practical guidance to researchers performing economic evaluations of cancer screening.

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

2. Comparisons between programs

Full economic evaluations are always comparative. No intervention is ever economically attractive or unattractive in isolation. Each is always more or less so in relation to some other program, even if the other program is "no intervention."

What programs should be compared? An economic evaluation ideally includes all potential, mutually exclusive programs for a defined population, including, of course, the option of not screening at all. In cancer screening, the two key variables are screening modality and screening frequency. Thus, an "ideal" economic evaluation of colorectal cancer screening would involve the comparison of all methods of stool occult blood testing, sigmoidoscopy, colonoscopy, and barium enema, singly and in every combination over all possible screening intervals. This is clearly a heroic task, though some analysts have compared up to 40 alternative screening programs [2]. Study design and modeling limitations (tractability) always limit the consideration of therapeutic alternatives. Economic evaluations carried out in parallel with randomized controlled trials, for example, may be limited to the examination of only two or three alternatives. There is no widely accepted formal approach to determining which subset of possible alternatives should be examined. One set of guidelines recommends that the least expensive, most effective, and most widely used strategies should be evaluated [3]. The policy recommendations of specialty groups like the American Cancer Society should also be considered, since they represent reasoned judgments about combinations of screening modality and frequency that will be economically and clinically attractive. The "no screening" alternative offers a useful benchmark that facilitates meaningful comparison across programs, and should be evaluated whenever feasible.

3. Study architecture: modeling and clinical trials

There are two fundamentally different approaches to gathering and combining the cost and health benefit data in an economic evaluation. The first approach is to link an economic evaluation directly to a randomized controlled trial (RCT) designed to evaluate the effectiveness of screening. In this "piggyback" approach [4], the costs and clinical consequences of screening are gathered simultaneously, and the resource implications of gaining extra health benefit are computed directly from trial data at completion [4,5]. This approach has not yet seen wide application in the evaluation of cancer screening programs, though there is a trend to incorporate economic analyses into the design of new clinical trials [4–6].

One advantage of linking an economic evaluation to an RCT is that cost and effectiveness data are at least theoretically more valid, subject as they are

to the blinding, randomization, and explicit definition of outcomes that minimize the potential for bias in RCTs. In addition, comprehensive cost data are usually relatively easy to gather within the context of a clinical trial. On the other hand, the representativeness of cost data gathered within a trial is often difficult to determine [7]. It is not always straightforward to separate the costs associated with trial infrastructure from costs associated with the delivery of health care. In addition, clinical practice associated with screening and treatment within a trial is highly stylized and may bear little relation to community practice [5]. Thus, it may be difficult to be certain that the results of an economic evaluation carried out within the context of a trial are applicable to a given clinical setting.

Clinical trials usually run for periods of time shorter than the optimal time horizon (the period of time covered by the analysis). Thus, economic evaluations that are piggybacked onto trials often require additional modeling work to project costs and clinical outcomes into the future, as well as to assess the degree of uncertainty associated with an evaluation [8,9]. A final problem, of course, is that one can only examine a very limited number of programs in a trial.

The alternate approach, which has been much more widely used in analyses of cancer screening to date, is to compare screening alternatives using decision-analytic or simulation models [10–15]. In this approach, cost data and health data are gathered from disparate (usually secondary) sources and incorporated into a common model. In some modeling exercises, costs and/or quality-of-life effects may actually be measured, but effectiveness data usually are not. The advantages and disadvantages are the obverse of evaluations carried out in the context of a trial. Modeling allows the evaluation of many more programs at more intervals, but the comparisons between programs may be subject to greater bias. Modeling is useful when the results of clinical trials are conflicting [14], when clinical trials have not yet been performed [15], or when the time horizon of clinical trials is limited [16]. In addition, modeling studies are the only means by which the population effects, both clinical and economic, of alternative screening strategies can be estimated.

Both approaches are subject to certain limitations. We believe that the validity of the evaluation is less affected by study architecture than by validity of the data, though the two may be related. It is of particular importance that the effectiveness of screening in reducing cancer mortality be accurately represented in the analysis. When clinical trials of screening yield conflicting results [14], or when estimates of efficacy are derived from biological models [2] or less rigorous clinical data (e.g., comparison of cohort studies [15]), economic evaluations may be useful in sketching out the potential effects of screening, but will remain less than definitive. As always, the validity of a cost-effectiveness analysis is contingent on the validity of the effectiveness data.

4. Costs

The cost of any component of screening is the value of the resources required to deliver the service. This value is theoretically straightforward, but measuring costs is often difficult. Charges associated with health services are much easier, in general, to obtain than costs and are frequently used in economic analyses [17]. Charges, however, do not bear a consistent relation to costs, since individual services may either be profitable (charges > costs) or subsidize other services (costs > charges) [18].

Cost and charge data may be derived from single institutions (e.g., hospitals, managed care or capitated health care organizations), insurers, or government (e.g., provincial Ministries of Health, Medicare, Medicaid). Ideally, economic evaluations should utilize actual costs drawn from a large and representative sample. In practice, there is often a tradeoff between accuracy of costing and the representativeness of the sample from which costs are derived [15,19].

5. Cost categories

There are many different types of cost, and not all should be included in every analysis. The major categories of cost are direct, indirect, and intangible costs. Direct costs are expenditures that are induced or averted by an intervention. Direct costs may be related to provision of a medical service (direct health care costs). Examples here would be the costs associated with the screening maneuver, the confirmatory tests, and the definitive cancer treatment (see table 1). However, patients must travel to their appointments and may incur lodging expenses or expenses related to special diets, clothing, or prostheses. These costs are direct in that they are expenditures induced by the intervention, but are nonmedical, and hence are termed direct nonmedical or non-health care costs.

Cancer and interventions to screen or treat cancer may result in productivity loss, as well as direct costs. Individuals who are ill, or die prematurely, will not be able to work and contribute productively to the economy. The concept of indirect cost, then, refers to the productivity loss caused by disease-related morbidity and premature mortality.

Productivity losses caused by disease represent real costs, but there is no universally agreed upon method for valuing these losses [20–23]. The practical difficulty of measuring productivity loss, as well as methodologic disagreement about exactly how to do so, accounts for the fact that few analyses of cancer screening include indirect costs.

Intangible cost is the value of pain, suffering, grief, and other nonfinancial outcomes associated with illness. Though it is possible to express the "cost" of pain and suffering by attaching a monetary value (as is done in cost–benefit analysis), it is more common to express intangible costs as health outcomes.

Table 1. Costs in Cancer Screening Programs

	Fahs et al. [10]	Eddy et al. [11]	Shimbo et al. [75]	Krahn et al. [15]	de Koning et al. [49]	Carter et al. [19]
Direct medical costs						
Program start-up costs	—	—	—	—	×	×
Advertising/invitation to screening	—	—	—	—	×	×
Screening examination	×	×	×	×	×	×
Confirmatory tests	?	×	×	×	×	×
Cancer treatment	×	×	×	×	×	×
Follow-up of false-positive, weakly positive, or suspicious tests	?	×	—	—	?	?
Regular follow-up posttreatment	×	—	×	×	×	—
Initial and follow-up care for treatment complications	×	×	×	×	—	—
Initial and follow-up treatment of recurrence of progression	×	×	?	×	—	—
Initial and follow-up costs of treatment of advanced or metastatic disease	×	×	?	×	×	×
Terminal costs of disease	?	×	×	×	—	—
Cost of treating other conditions during additional years of life	—	×	—	—	—	—
Direct nonmedical costs						
Transportation	—	—	—	—	—	—
Home care	—	—	—	—	—	—
Indirect costs						
Productivity loss due to illness (morbidity)	—	—	—	—	—	—
Productivity loss due to premature death (mortality)	—	—	—	—	—	—
Productivity loss associated with travel, waiting, and hospitalization	—	—	—	—	—	—
PERSPECTIVE	Societal	Not stated	Not stated	Third-party payer	Not stated	Not stated
COSTS/CHARGES	Charges	Charges	Charges	Costs	Costs/charges	Costs
SOURCE OF COSTS/CHARGES	Medicare	Medicare, LI[b]	PUB[a]	LI[b]	PROG[c], LI[b]	PROG[c], LI[b]
DISCOUNTING	5%	5%	5%	5%	5%	5%

[a] Costs derived from local institutions(s).
[b] Costs derived from published estimates.
[c] Costs are local program costs.

Thus, intangible costs are usually represented in the denominator of cost-effectiveness or cost-utility ratios.

6. Costs in cancer screening programs

Table 1 lists some of the costs that should be considered in the evaluation of a screening program, using selected studies as examples. In general, most published analyses gather relatively complete costs relating to the screening method(s) as well as treatment for those found to have cancer. In addition, most analyses attempt in some way to measure the costs of ongoing care for those found to have cancer, including the costs of treating recurrent, progressive, or advanced disease.

The initial costs of setting up a cancer screening program, which include capital, equipment, training costs, and advertising, are almost always overlooked, because analyses often evaluate hypothetical or already functioning programs. Failure to include these costs will result in a bias toward the screening program, since the unit cost per positive screen will be falsely low.

In addition, the economic consequences of false-positive or indeterminate tests are often omitted. For example, an elevated prostate-specific antigen level will almost always trigger follow-up visits and repeat transrectal biopsies, even if the inital biopsy is negative. Clinicians may be concerned that the patient is still at risk for cancer and may believe that the initial biopsy "missed" the tumor. For many screening methods, false positives exceed true positives, so the economic consequences of altered clinical behavior after the initial test are potentially very important and should be included in a full evaluation. Again, the consequence of omitting this cost component is to bias the analysis toward the screening program by underestimating the full cost of screening.

The accuracy of estimates of ongoing treatment costs varies. Modeling studies, in particular, often assign treatment costs by clinical stage, irrespective of the history of prior treatment [2]. In practice, treatment decisions will incorporate information about prior treatment. Treatment for a stage C prostate cancer, for example, may depend on whether the patient has already received surgical treatment or radiotherapy. Empirical data about the lifetime stream of cancer treatment costs or data gathered within the context of an RCT are more plausible estimates of true treatment costs than those that predict initial and follow-up costs from cancer stage.

Direct, nonmedical costs are rarely included in published analyses, even those that explicitly adopt a societal perspective [10]. Because the magnitude of these costs is small relative to the costs of screening and treatment, and because the cost of gathering these data is often high, requiring patient surveys, this omission is rarely important.

7. Perspective of the analysis

Which costs should be included within the analysis is determined by the perspective of the analysis. Economic evaluations may adopt the perspective of society, government, governmental or other third-party payer, hospital, department, patient, or other individual health care provider. The societal perspective is the broadest perspective and takes into account all costs, irrespective of their distribution. Analyses that adopt a societal perspective should include, in addition to direct medical costs, all productivity losses, direct nonmedical costs, and costs to individuals.

The most frequently utilized perspective is that of the payer (government or other third party), which excludes costs borne by individuals for direct health care and insurance premiums, as well as excluding direct nonhealth costs and indirect costs. This perspective is arguably the most appropriate, since decision makers within health care organizations will attempt to maximize health outputs subject to their budgetary constraint [24,25]. The most comprehensive approach is to carry out the analysis from all perspectives, including societal and individual perspectives. If the optimal decision differs according to the perspective of the analysis, this will be brought into sharp relief by the analysis and can be taken explicitly into account by the decision maker(s) [1].

8. Health effects

The way in which health effects are measured is the chief determinant of the type of economic analysis. The simplest way of dealing with health effects is to ignore them. This is perfectly legitimate in the rare situation where two programs are of equal efficacy (cost-minimization study). Similarly, a simple tabulation of costs may show that one program is more economically attractive than another, if it is less costly and is known to produce greater health benefit (cost comparison).

Because of the complexity of cancer screening, partial evaluations, as described above, are rarely sufficient. It is usually necessary to measure health effects as well as costs in a full evaluation. Health effects may be measured in natural units, e.g., cancers detected or lives saved, in which case the analysis is a cost-effectiveness analysis. A limitation of this approach is that the denominator is expressed using different measurement units, so comparison across clinical domains is difficult.

An alternative and more general approach is to convert health outputs into standardized units, e.g., quality-adjusted life years [26,27], or healthy year equivalents [28]. This type of analysis is known as cost-utility analysis. The chief virtue of this approach is that it allows comparison between programs with dissimilar health outputs, both within and across clinical domains. Thus, it is feasible to compare the economic attractiveness of screening

mammography with breast cancer treatment, as well as in-home hemodialysis or bone-marrow transplantation for acute leukemia, by determining the incremental cost of producing a standardized health unit in each program.

The most widely used unit of health output is the quality-adjusted life year. In this conceptualization, health has two dimensions, namely, length of life and quality of life. Utility, a measure of patient preference, is used as a weight to adjust length of life for varying quality and produce a standardized unit [26,27,29]. The healthy year equivalent [28] has been proposed as a theoretically preferable measure, but has not yet been widely applied because it requires additional measurement and necessitates severe modeling constraints.

The final approach to characterizing health benefits is to express their value in monetary terms. This dramatically simplifies interpretation of the analysis, since costs and health effects are expressed in the same units (i.e., dollars). A net cost saving or net cost increase suggests that a program is economically attractive or unattractive. This approach has always had more attraction for economists than health care decision makers, since it requires the analyst to value health and life in monetary terms, a step that often engenders distrust among consumers of the research [30]. As Phelps and Mushlin [31] have pointed out, however, cost effectiveness and cost utility also require implicit valuation of health outcomes in monetary terms if they are used to direct policy decisions. Thus, under most circumstances, all three forms of full economic evaluation are equivalent (though not equally informative), despite differences in reporting style.

9. Health effects in cancer screening programs

Screening programs generate health effects that start with the screen and may persist over decades. The screen itself may be innocuous (PSA) or it may be uncomfortable and invasive (colonoscopy, transrectal or transvaginal ultrasound). Though the effect per individual will be small, all screenees will be affected, as opposed to the few who will benefit from screening.

The psychological effects of screening must also be considered. Those who have cancer identified early through screening will all live longer with the diagnosis, though only some will live longer. All screenees confront their own mortality, though this effect is not necessarily negative [32–34]. Patients with false-positive or equivocal tests may live with a heightened fear of cancer, and may seek or be given more intensive, potentially invasive, medical care. Screening may also reassure those who fear cancer, though cancer anxiety may initially be generated by the publicity surrounding screening programs [35]. The effect of the initial screen on rescreens must be considered. Psychological effects of the initial screen may reduce compliance with subsequent screens [33,36]. Cancer treatment is usually invasive, involving surgery or radiation, the health effects of which are, all other things being equal, negative in the short term [37–39]. Treatment may be disfiguring (mastectomy), involve

changes in body image and self-perception (impotence post radical prostatectomy, colostomy post bowel resection, infertility post hysterectomy), and impair occupational, social, and sexual functioning (urinary and bowel incontinence, impotence) [40–43]. Screening may prevent more men from living with disease recurrence, cancer progression [44,45], advanced disease, and preterminal disease (fatigue, depression, inanition, cachexia, pain), so these effects must be considered [41,46–48]. Treatment for more advanced disease (chemotherapy, hormonal therapy, radiotherapy) must also be considered. Finally, the effects of screening on cancer mortality must be included.

As table 2 illustrates, few published economic evaluations of cancer screening evaluate the full range of screening-related health effects. Nearly all analyses are cost-effectiveness analyses, whose final result is expressed as an incremental cost per life year gained. The psychological effects of screening and the quality-of-life effects of treatment, treatment complications, and disease progression are omitted from this type of analysis.

At least two screening studies have measured quality of life. In an evaluation of breast cancer screening in the Netherlands [49], measurement of quality of life did not have a large impact on the outcome of the analysis. The modest negative effects on health of screening, treatment, and living longer with a cancer diagnosis were almost exactly offset by the reduction in late cancer morbidity induced by cancer screening. Thus, the quality-adjusted life expectancy gain was very close to the unadjusted life expectancy gain. On the other hand, our evaluation of prostate cancer screening demonstrated that the negative effects of prostate treatment alone were sufficient to offset the mortality gain afforded by screening for prostate cancer [15]. Predicting the net effect of quality of life on an analysis is difficult without actually measuring these effects and incorporating them into the analysis.

10. Time horizon and discounting

The annual costs and benefits of health programs usually vary with time. Startup costs are often high, but maintenance costs are lower. As time goes by, annual program savings may exceed program costs. Preventive programs with long time horizons (e.g., cancer screening) often take years or even decades to achieve positive health benefits. Alternatively, health benefits may diminish with time (behavior modification, disease education). Choice of an inappropriately short time horizon can bias results of economic analyses by not capturing economic or health consequences of the program that extend beyond the horizon of the analysis. A minimum time horizon for evaluating a cancer screening program is the period during which screening can be expected to exert some effect on the disease-specific mortality rate. Shorter time horizons will bias the analysis against screening by not capturing all the potential screening benefit.

Economic costs and health effects that occur in the future are convention-

Table 2. Health effects in cancer screening programs

	Fahs et al. [10]	Eddy et al. [11]	Shimbo et al. [75]	Krahn et al. [15]	de Koning et al. [49]	Carter et al. [19]
DESIGN	Cost-effectiveness	Cost-effectiveness	Cost-effectiveness	Cost-utility	Cost-utility	Cost-effectiveness
Effects of screen itself (physical discomfort, pain)	—	—	—	—	×	—
Early labeling, cancer anxiety in cancer patients	—	—	—	—	×	—
Labeling, cancer anxiety in false positives or indeterminate tests	—	—	—	—	—	—
Reassurance that no cancer is present	—	—	—	—	—	—
Discomfort, pain of confirmatory tests	—	—	—	—	×	—
Effects of cancer treatment	—	—	—	×	×	—
Effects of treatment complications	—	—	—	×	—	—
Effects of disease recurrence/progression	—	—	—	×	—	—
Effects of treatment for disease progression/recurrence	—	—	—	×	×	—
Advanced or metastatic disease	—	—	—	×	×	—
Treatment of advanced or metastatic disease	—	—	—	×	×	—
Life years gained	×	×	×	×	×	—
Method of utility assessment	—	—	—	Time-tradeoff	Rating scale	

ally assigned a lower value (discounted) than those that occur in the present [50]. Future costs are discounted because, given the choice, most individuals would prefer to spend the same amount later or receive a similar amount sooner, even after adjustment for inflation (positive time preference). In addition, resources not used can be invested: $1000 invested at 10% will be $1100 one year from now, so the present and future value of $1000 are quite different (opportunity cost). Thus, there is general agreement that costs should be discounted [50–57]. It is also generally [57], though not universally [58], accepted that future health benefits should also be discounted. Future health and economic effects of health programs are usually devalued at a common rate, most frequently 5% [50,59]. The choice of a discount rate is less important in economic evaluations with a short time horizon, but often very important in programs whose costs are immediate but whose benefits take longer to accrue.

11. Evaluating uncertainty: sensitivity analysis

Sensitivity analysis is the main tool used to evaluate whether the qualitative conclusion reached in the analysis is robust to the uncertainties in the model. The simplest form of sensitivity analysis involves varying a single parameter across the range of uncertainty to determine whether the analytic result changes. A more complex form involves the simultaneous evaluation of uncertainty in two or three variables. "Structural" sensitivity analysis involves changes in model structure to determine whether modeling assumptions are creating bias. Analysis of extremes is another frequently employed technique: by varying multiple parameters simultaneously, the model is systematically biased toward and then against a given program to see if the baseline analytic result is robust [60]. Finally, probabilistic sensitivity analysis can simultaneously evaluate the effects of uncertainty in many input parameters [61–63].

Sensitivity analysis is vitally important in economic evaluation precisely because so many assumptions are required. Despite their complexity, published analyses of cancer screening have not, in general, included very complete sensitivity analyses. The most common practice is to report one-way sensitivity analyses for a very selected group of model parameters. Multiway analyses, analyses of extremes, and structural and probabilistic sensitivity analysis are rare. Thus, it is often extremely difficult for a consumer of published analyses to know how much uncertainty is attached to a reported result.

12. Interpreting published analyses

12.1. From internal to external validity

Economic evaluations of breast cancer screening have reported incremental costs per year of life saved ranging from $3000 to $84,000 [64]. Why is there

such extreme variation in reported results, and how confident can we be in applying the results of such research in our own clinical or health care context?

First, apparent differences in reported results may be accounted for by methodologic differences. The perspective of the analysis, inclusion/omission of selected costs, the chosen discount rate, the target year of the analysis (inflation), the time horizon of the analysis, the way in which clinical efficacy is modeled or expressed, and whether average or incremental cost-effectiveness ratios are reported [25] may all affect the reported results.

In addition, consumers of the literature must keep in mind that economic evaluations consider extremely specific clinical strategies, for specific populations, within a given national health care system. Comparing screening mammography to no screening and to regular breast self-examination will yield wildly different cost-effectiveness ratios. Different screening intervals [49], high- and low-risk populations [15,65], young or old populations [10,14] and the efficiency of the health care system [66] may all dramatically affect cost-effectiveness ratios.

Thus, the usual concerns about the generalizability of published clinical research to one's own setting [67,68] apply to economic research as well, but perhaps with additional force, since generalizability depends not only on clinical but also on economic similarity between study and target populations. The economic attractiveness of a clinical strategy cannot be generalized to alternate populations, alternate health care systems, varying screening intervals, or combinations of screening tests.

12.2. Decision criteria

When cancer screening programs cost more than an alternative and offer the potential for clinical harm [15], the alternative is clearly preferred. When screening is cost saving and produces net health benefit, the program ought to be implemented. Unfortunately, cancer screening usually produces a net health increase at an increase in net cost. This situation is more difficult to interpret. Whether the screening program ought to be implemented depends on how much society values (is willing to pay for) a standardized unit of health. This is not a question of fact, but of value or preference. How a "society" values something depends, of course, on the society and who within that society one asks. The economic attractiveness of health care interventions that society is paying for at present (e.g., center hemodialysis, hypertension treatment, coronary revascularization) offers some, albeit limited, guidance [69]. Laupacis et al. [70] offer suggested thresholds of economic attractiveness (<$20,000/QALY: strong evidence for adoption; $20,000–$100,000/QALY: moderate evidence for adoption; >$100,000/QALY: weak evidence for adoption), but these thresholds are by no means uncontroversial or universally accepted.

13. Economic evaluation: summary and caveats

If economic evaluation is to be valuable in policy formulation, its limitations must be understood. First, economic evaluation is not prescriptive, but rather a first (albeit important) step in the evaluation of a new health care technology. The distribution of costs and health benefits (who gains and who loses) and availability (matching resources to locations where they are accessible to those who require them) must also be considered [25].

An additional reason why the results of efficiency evaluation should not be blindly applied is that the methodology itself involves a standard theoretical framework that does not always validly represent the values of individuals or society. Two aspects of efficiency evaluation are particularly germane to cancer screening.

The first is the way in which preferences for health states are represented in economic models. Empirical studies have consistently shown that different individuals may assign very different values to standardized health states (e.g., mastectomy, impotence or incontinence). Yet a common approach among analysts is to express population preferences in models using a measure of central tendency, usually a mean utility value. The analytic result produced by this assumption may well be the correct one for most individuals, but the unthinking application of the policy suggested by the analysis may do individuals harm. Consumers of the economic evaluation literature should exercise particular caution when sensitivity analyses demonstrate that the optimal choice is contingent on patient preferences for health outcomes.

Finally, there is the issue of risk. Cost-effectiveness analysis and its relatives produce an estimate of expected cost and expected health loss or gain per person that does not explictly account for risks faced by each patient [1]. For example, program A and program B each result in a gain of one quality-adjusted day per person relative to program C. In program A, each individual gains one day. In program B, most individuals are unaffected, but a very few are prevented from dying prematurely (30 quality-adjusted life years lost). Though many individuals will value program A and program B differently (most will prefer B), conventional economic evaluation does not. Thus, programs that decrease risk of very adverse outcomes, like (for the most part) cancer screening, are undervalued relative to programs that are risk neutral or risk increasing. There is no widely accepted formal way of adjusting for risk [50], but this is an important limitation of economic evaluation, and may partly account for some of the difficulties associated with translating health policy into clinical practice [71].

References

1. Drummond MF, Stoddart GL, Torrance GW. 1980. Principles of Economic Appraisal in Health Care. Oxford: Oxford University Press.

2. Eddy DM, Nugent FW, Eddy JF, et al. 1987. Screening for colorectal cancer in a high-risk population. Gastroenterology 92:682–692.
3. Task Force on Principles for Economic Analysis of Health Care Technology. 1995. Economic analysis of health care technology. Ann Intern Med 122:61–70.
4. Bonsel GJ, Rutten FFH, Uyl-de Groot CA. 1993. Economic evaluation alongside cancer trials: methodological and practical aspects. Eur J Cancer 29a:s10–s14.
5. Drummond MF, Davies L. 1991. Economic analysis alongside clinical trials. Revisiting the methodological issues. Int J Technol Assessment Health Care 7:561–573.
6. Kramer BS, Gohagan J, Prorok PC, et al. 1993. A National Cancer Institute sponsored screening trial for prostatic, lung, colorectal, and ovarian cancers. Cancer 71:589–593.
7. Evans RG, Robinson GC. 1980. Surgical day care: measurement of the economic payoff. CMAJ 123:873–880.
8. Bonsel GJ, Klompmaker IJ, van't Veer F, et al. 1990. Use of prognostic models for assessment of value of liver transplantation. Lancet 335:493–497.
9. van Hout BA, Bonsel GJ, Habbema JDF, et al. 1993. Heart transplantation in the Netherlands. J Health Econ 12:73–93.
10. Fahs MC, Mandelblatt J, Schechter C, et al. 1992. Cost effectiveness of cervical cancer screening for the elderly. Ann Intern Med 117:520–527.
11. Eddy DM. 1990. Screening for colorectal cancer. Ann Intern Med 113:373–384.
12. Eddy DM. 1989. Screening for breast cancer. Ann Intern Med 111:389–399.
13. Barry MJ, Mulley AG, Richter JM. 1987. Effect of workup strategy on the cost-effectiveness of fecal occult blood screening for colorectal cancer. Gastroenterology 93:301–310.
14. Eddy DM. 1988. The value of mammography screening in women under age 50 years. JAMA 259:1512–1519.
15. Krahn MD, Mahoney JE, Eckman MH, et al. 1994. Screening for prostate cancer: a decision analytic view. JAMA 272:781–786.
16. Mushlin AI, Fintor L. 1992. Is screening for breast cancer cost-effective. Cancer 69:1957–1962.
17. Finkler SA. 1982. The distinction between cost and charges. Ann Intern Med 96:102–109.
18. Eisenberg JM. 1989. Clinical economics: a guide to the economic analysis of clinical practices. JAMA 262:2879–2886.
19. Carter R, Glasziou P, van Ootmarssen G, et al. 1993. Cost-effectiveness of mammographic screening in Australia. Aust J Public Health 17:42–50.
20. Cooper BS, Rice DP. 1976. The economic costs of illness revisited. Soc Security Bull 39:21–36.
21. Rice DP, Hodgson TA, Kopstein AN. 1985. The economic costs of illness: a replication and update. Health Care Fin Rev 7:61–80.
22. O'Brien B, Viramontes JL. 1994. Willingness to pay: a valid and reliable measure of health state preference. Med Decision Making 14:289–297.
23. Robinson JC. 1986. Philosophical origins of the economic valuation of life. Milbank Q 64:133–155.
24. Weinstein MC. 1990. Principles of cost-effective resource allocation in health care organizations. Int J Technol Assess Health Care 6:93–103.
25. Detsky AS, Naglie IG. 1990. A clinician's guide to cost-effectiveness analysis. Ann Intern Med 113:147–154.
26. Torrance GW, Feeny D. 1989. Utilities and quality-adjusted life years. Int J Technol Assess Health Care 5:559–575.
27. Shepard DS, Weinstein MC. 1980. Utility functions for life years and health status. Oper Res 28:206–224.
28. Mehrez A, Gafni A. 1989. Quality adjusted life years, utility theory, and healthy year equivalents. Med Decision Making 9:142–149.
29. Torrance GW. 1987. Utility approach to measuring health-related quality of life. J Chron Dis 40:593–600.
30. Weinstein M, Fineberg H. 1980. Clinical Decision Analysis. Philadelphia: WB Saunders.
31. Phelps C, Mushlin AI. 1991. On the (near) equivalence of cost-effectiveness and cost-benefit analyses. Int J Technol Assess Care 7:12–21.

32. Lerman C, Trock B, Rimer BK, et al. 1991. Psychological and behavioral implications of abnormal mammograms. Ann Intern Med 114:657–661.
33. Lerman C, Rimer BK, Trock B, et al. 1990. Factors associated with repeat adherence to breast cancer screening. Prev Med 19:279–290.
34. Reelick NF, de Haes WF, Schuurman JH. 1984. Psychological side effects of the mass screening on cervical cancer. Soc Sci Med 18:1089–1093.
35. Wilkinson GB, Meacham AR. 1991. Prostate cancer awareness week: how we did it. Urol Nursing 11:19–20.
36. Greenwald HP, Becker SW, Nevitt MC. 1978. Delay and noncompliance in cancer detection: a behavioral perspective for health planners. Milbank Mem Fund Q Health Soc 56:212–230.
37. Patterson DE, Zincke H. 1984. Perioperative complications of pelvic lymphadenectomy and radical retropubic prostatectomy in 150 patients. Urology 23:243–246.
38. Morse RM, Spirnak P, Resnick MI. 1988. Iatrogenic colon and rectal injuries associated with urological intervention: report of 14 patients. J Urol 140:101–103.
39. Pilepich MV, Perez CA, Walz BJ, et al. 1981. Complications of definitive radiotherapy for carcinoma of the prostate. Int J Radiat Oncol Biol Phys 7:1341–1348.
40. Herr HW. 1994. Quality of life of incontinent men after radical prostatectomy. J Urol 151:652–654.
41. Kornblith AB, Herr HW, Ofman US, et al. 1994. Quality of life of patients with prostate cancer and their spouses. The value of a data base in clinical care. Cancer 73:2791–2802.
42. Schover LR. 1993. Sexual rehabilitation after treatment for prostate cancer. Cancer 71 (3 Suppl):1024–1030.
43. Walsh PC, Donker PJ. 1982. Impotence following radical prostatectomy: insight into etiology and prevention. J Urol 128: 492–497.
44. Schellhammer PF, el-Mahdi AM. 1990. Local failure and related complications after definitive treatment of carcinoma of the prostate by irradiation or surgery. Urol Clin North Am 17:835–851.
45. Schellhammer PF, Whitmore RB, Kuban DA, et al. 1989. Morbidity and mortality of local failure after definitive therapy for prostate cancer. J Urol 141:567–571.
46. Fossa SD, Aaronson N, Calais da Silva F, et al. 1989. Quality of life in patients with muscle-infiltrating bladder cancer and hormone-resistant prostatic cancer. Eur Urol 16:335–339.
47. Fossa SD, Aaronson N, de Voogt HJ, et al. 1990. Assessment of quality of life and subjective response criteria in patients with prostatic cancer. In EORTC Genitourinary Group Monograph 7: Prostate Cancer and Testicular Cancer. New York: Wiley-Liss, p. 199.
48. Fossa SD, Aaronson NK, Newling D, et al. 1990. Quality of life and treatment of hormone resistant metastatic prostatic cancer. Eur J Cancer 26:1133–1136.
49. de Koning HJ, van Ineveld BM, van Ootmarssen GJ. 1991. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J Cancer 49:531–537.
50. Krahn M, Gafni A. 1993. Discounting in the evaluation of health care interventions. Med Care 31:403–418.
51. Olson M, Bailey MJ. 1981. Positive time preference. J Polit Econ 81:1–25.
52. Fisher I. 1935. The Theory of Interest. London: Macmillan.
53. Baumol WJ. 1968. On the social rate of discount. Am Econ Rev 58:788–802.
54. Feldstein MS. 1964. The social time preference discount rate in cost benefit analysis. Econ J 74:360–379.
55. Goodin RE. 1982. Discounting discounting. J Public Policy 2:53–72.
56. Pigou AC. 1920. The Economics of Welfare. London: Macmillan.
57. Keeler EB, Cretin S. 1983. Discounting of life saving and other non-monetary benefits. Management Sci 29:300–306.
58. Hillman AL, Kim M. 1994. Decision making in health care: discounting of life years revisited. PharmacoEcon 7:198–205.
59. Weinstein MC, Stason WB. 1977. Foundations of cost-effectiveness analysis for health and medical practices. N Engl J Med 296:716–721.

60. Briggs A, Sculpher M, Buxton M. 1994. Uncertainty in the economic evaluation of health care technologies: the role of sensitivity analysis. Health Econ Ser 3:95–104.
61. Critchfield GC, Willard KE. 1986. Probabilistic analysis of decision trees using Monte Carlo simulation. Med Decision Making 6:85–92.
62. Willard KE, Critchfield GC. 1986. Probabilistic analysis of decision trees using symbolic algebra. Med Decision Making 6:93–100.
63. Doubilet P, Begg CB, Weinstein MC, et al. 1985. Probabilistic sensitivity analysis using Monte Carlo simulation. Med Decision Making 5:157–177.
64. Brown ML, Fintor L. 1993. Cost-effectiveness of breast cancer screening: preliminary results of a systematic review of the literature. Breast Cancer Res Treatment 25:113–118.
65. Nutting PA, Calonge BN, Iverson DC, et al. 1994. The danger of applying uniform clinical policies across populations: the case of breast cancer in American Indians. Am J Public Health 84:1631–1636.
66. van Ineveld BM, Oortmarssen GJ, de Koning HJ, et al. 1993. How cost-effective is breast cancer screening in different EC countries? Eur J Cancer 29A:1663–1668.
67. Sackett DL, Haynes RB, Guyatt GH, et al. 1991. Clinical Epidemiology: A Basic Science for Clinical Medicine, 2nd ed. Boston/Toronto/London: Little, Brown and Company.
68. Guyatt GH, Sackett DL, Cook DJ. 1994. User's Guides to the Medical Literature 11. How to use an article about therapy of prevention B. What were the results and will they help me in caring for my patients? JAMA 271:59–63.
69. Drummond M, Torrance G, Mason J. 1993. Cost-effectiveness league tables: more harm than good? Soc Sci Med 37:33–40.
70. Laupacis A, Feeny D, Detsky AS, et al. 1993. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. CMAJ 148:921–924.
71. Asch DA, Hershey JC. 1995. Why some health policies don't make sense at the bedside. Ann Intern Med 122:846–850.
72. Ben-Zion U, Gafni A. 1983. Evaluation of public investment in health care: is the risk irrelevant? J Health Econ 2:161–165.
73. Gafni A, Torrance GW. 1984. Risk attitude and time preference in health. Management Sci 30:440–451.
74. Arrow KJ, Lind RC. 1970. Uncertainty and the evaluation of public investment. Am Econ Rev 60:364–378.
75. Shimbo T, Glick HA, Eisenberg JM. 1994. Cost-effectiveness analysis of strategies for colorectal cancer screening in Japan. Int J Technol Assess Care 10:359–375.

4. Screening for cervical cancer

Matti Hakama

1. Benefits and harms of screening

Screening for cancer involves the identification of preclinical disease by a relatively simple test. The objective of screening is to reduce the risk of death, i.e., mortality from cancer among subjects subjected to screening. For cervical cancer, the screening test is aimed at detection of preinvasive lesions. Therefore, reduction in the incidence of invasive disease results from screening for cervical cancer, and an indicator for the effect is change of incidence in time before and after application of the screening test, or difference in incidence between those subjected to screening and those not subjected to screening.

Screening may have benefits other than an effect on incidence and mortality. If the treatment of disease detected at screening is less invasive or less radical, or results in less morbidity than that of clinically detected disease, then the quality of life of the screened population is improved. Correct negative results also have a beneficial effect in that they reassure people without the disease.

Because the aim of cervical screening is to provide a preinvasive diagnosis of disease, there is a prolonged period of morbidity — from the time of diagnosis at screening to the hypothetical time at which a clinical diagnosis would have been made had the patient not been screened. This lead time, while a prerequisite of effective screening, is an adverse effect because of the prolongation of anxiety and morbidity due to diagnosis and treatment of cancer.

Cases detected at screening are confirmed by standard clinical diagnostic methods. Many such cases are borderline abnormalities, some of which would progress to clinical disease and some of which would not, even if left untreated. The diagnosis of carcinoma in situ and severe dysplasia results in an invasive treatment. A proportion of these lesions would not have progressed to clinical disease during the woman's lifespan [1]. Any screening program will disclose such abnormalities, which are indistinguishable from the truly abnormal cases that will progress into the clinical phase in the absence of early treatment. One of the adverse effects of screening is therefore the consequent treatment of screenees with such lesions. This results in anxiety and morbidity. Also, false-

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

positive screening results cause anxiety and sometimes morbidity even if later confirmed to be benign. Conversely, a false-negative result is falsely reassuring. If it results in postponement of clinical diagnosis and worsens the outcome of treatment, screening is disadvantageous.

Many screening programs involve expensive techniques and are applied to large populations. The total budget required for health services is thus likely to increase if a screening program is adopted.

A decision about whether to screen requires weighing the beneficial and harmful effects. As is general in medicine, value judgments are involved in such weighting, so ethical issues are also closely related to screening.

2. Effectiveness

The effectiveness of screening for cervical cancer in terms of reduced incidence has never been demonstrated in a randomized preventive trial. There is, however, nonexperimental evidence on reduction in incidence of invasive disease. Canada was the pioneer in screening for cervical cancer [1,2]. The largest of the studies is the collaborative study coordinated by the International Agency for Research on Cancer [3], which showed that eradication of the disease is an unrealistic goal and that maximal protection after a negative smear is about 90%, which remains roughly the same during several years after the test (table 1). This conclusion is in accordance with the results of studies on the natural history of the disease, which have shown that most preinvasive lesions progress to frankly invasive cancer only over several years [1].

Screening is public health policy, and the success of the whole program must be assessed. The validity of the program depends on the screening test used, attendance, the screening interval, and the success of referral for diagnostic confirmation of cases found at screening.

Much of the information on the applicability of screening for cervical cancer as a public health policy stems from the organized programs practiced in the Nordic countries since the mid-1960s [4]. Most of the Nordic countries

Table 1. IARC collaborative study: reduction of incidence of invasive cervical cancer after a negative smear by interval between the screens and age [3,20]

	Screening interval			
	% reduction in incidence		Number of lifetime smears	
Age	3 yrs	5 yrs	3 yrs	5 yrs
20–64	91	84	15	9
25–64	90	82	13	8
30–64	85	77	12	7
35–64	78	70	10	6

have nationwide screening programs for cervical cancer that fulfill the general prerequisites of an organized program [5] and make it possible to follow up each woman for the occurrence of intraepithelial cervical neoplasia and for cervical cancer. The programs define the ages and the frequencies of screening, use personal invitations with times and places for screening, and give personal information about the results of screening even when the smear is negative. Screening for cervical cancer reduces the incidence of invasive disease and is applicable as public health policy, but a wide variation is seen, from highly effective programs to relatively poor ones.

In Finland every women aged 30–55 receives a personal invitation at regular intervals (every five years) to attend. In the invitation letter, she is given an appointment place and time. The result of the Pap test is also given by mail, independent of whether the result is normal, suspicious, or malignant. Approximately 1,400,000 woman-years follow-up of 400,000 women, with information on the actual (participants) or potential (nonrespondents) first screening, were analyzed by a cohort design [6]. Among these women were the first ones, under the national policy, to reach the first rescreening after the five-year interval. The protective effect, in terms of reduction in the incidence of invasive disease among the responders, was about 80% (table 2). It could be argued that the responders were a selected group of the target population. Usually, in a spontaneous screening program, the target population is strictly unknown or undefined. In the organized Finnish sytem, the target population was identified from the national population registry and received a written invitation. The risk of cervical cancer among the total target population may remain unchanged in spite of reduction in the risk of attenders. The ultimate effect depends on the rate of attendance and the risk of cancer among attenders and among nonresponders. The incidence of invasive cervical cancer among the target population in the Finnish study, responders and nonresponders combined, was 40% of that among the controls, showing, therefore, a 60% protective effect due to screening. This estimate was a weighted average of the reduced risk among the attenders and the greater risk among the nonattenders.

While an effect of the selection by attendance described above on the protective effect could be ruled out, there still was the problem of unbiased

Table 2. The Finnish organized mass screening program for cervical cancer in 1963–1971: incidence [6]

	Age						
Screening status	30–34	35–39	40–44	45–49	50–54	55–59	Cumulative risk
Invitees							
Attenders	2	2	5	10	12	12	215
Nonattenders	19	34	37	82	68	26	1230
Noninvitees	7	20	35	46	47	46	1005

choice of controls for the target population. When a public health policy is run, the service is provided to everybody, and no control population remains. The Finnish study used the total Finnish female population as controls and used the incidence for the whole of Finland shortly before the start of the national program as the expected risk. It can be argued that there was a decreasing trend in the overall incidence already taking place before the start of the program. If so, then the control rates were too high, and the estimate of the protective effect was due to biased expected rates without any real effect due to screening. However, the lag between the control rates and screening rates was short and, if anything, there was an increasing trend in the overall incidence of cervical cancer in all the Nordic countries, including Finland, before the start of screening [7]. Such a trend was recently confirmed in Estonia, where no screening has been practiced [8]. With the liberalization of sexual mores, this result is what one would expect.

Within the organized programs, there are differences in cervical cancer screening policies between the Nordic countries. In Finland [4], Iceland [9], and Sweden [10], nationwide population-based organized programs have been in operation at least since the early 1970s, whereas only a few counties in Denmark, including the most populous ones, had organized screening programs [11,12]. The programs are run by voluntary cancer organizations in Finland and Iceland, and by the counties in Denmark and Sweden. The recommended age groups to be covered are 30–55 years in Finland (recently the program was extended up to 60 years), 25–69 years in Iceland, and 30–49 years in Sweden. The screening intervals recommended are two to three years in Iceland, four years in Sweden, and five years in Finland. In Denmark, the practice varies by county, but the National Board of Health recommendation is to have a smear every three years from the age of 23 to 59 and every five years from 60 to 75. In Norway [13,14], only 5% of the population was covered by an organized program. Cytological smears are, however, frequently taken outside the organized system, by private gynecologists and elsewhere. Such smears are taken more often than the smears in the organized programs in all the Nordic countries with the exception of Iceland.

In the Nordic countries, about 2500 new cases of cervical cancer were diagnosed annually before the screening programs were initiated. Since the early 1980s, the annual number of new cases has been about 1700. Denmark had a high incidence in the early period, its age-adjusted (world standard) incidence being about 30 per 10^5 women-years; in the other Nordic countries, the incidence was about 15, with somewhat increasing trends before the screening programs started. In the early 1980s, the rates ranged from 15 in Denmark downward [15].

There was a strong correlation between the extent of the organized screening program and changes in the incidence of invasive cervical cancer (table 3). The relative reduction in the risk was steepest in Finland and Sweden and intermediate in Denmark. In Norway the incidence rates of cervical cancer increased up to the 1970s. During the 15-year period from 1966–70 to 1981–85,

Table 3. Observed annual age-adjusted incidence rates (per 100,000 woman years) of invasive cervical cancer in the Nordic countries in selected time periods: data from the Nordic Cancer Registries [15,16]

	Observed				Predicted	
Country	1956–1960	1966–1970	1976–1980	1983–1987	1988–2002	2008–2012
Denmark	30	30	19	16	11	11
Finland	14	14	6	4	2	2
Iceland	16	26	9	13	10	9
Norway	15	17	17	13	9	8
Sweden	18	18	10	9	7	7

Table 4. Predicted numbers of deaths from cervical cancer in 1995 and 2015 in the Nordic countries assuming no screening (without) and the screening program as practiced in Finland (with) [18]

	Year			
	1995		2015	
Country	Without	With	Without	With
Denmark	530	90	530	40
Finland	290	70	310	30
Iceland	10	1	10	1
Norway	250	50	260	20
Sweden	470	100	480	40
Nordic	1550	310	1590	130

the incidence rates fell by 65% in Finland and 20% in Norway. The substantial decrease in incidence from the 1960s to the 1970s in Iceland is partly because prevalent microinvasive lesions were diagnosed during the first round of screening in the late 1960s more frequently than in the other Nordic countries. The rates in Iceland are subject to large random variation owing to the small population and relatively few cases of cancer.

The most substantial reduction in the risk of cervical cancer occurred in the age group 40–49 years of age [16], which probably came under the most intensive screening by the organized program. Again, the reduction was highest in Finland (80%) and lowest in Norway (50%). The rates somewhat increased at young ages, sharply decreased for the middle-aged, and were relatively stable for the elderly. The rates for women in Iceland were unstable owing to the small numbers.

Mortality was shown to follow closely the incidence trends in the Nordic countries [17]. The estimates for deaths from cervical cancers prevented (table 4) are also substantial. Compared to the hypothetical no-screening option, there was an estimated 76% reduction in the risk of death due to screening in 1995 if applied as in Finland [18].

It seems that the differences in the trends cannot be accounted for by the biology of the disease, but the most significant determinant of risk reduction is how well the program is organized. A comparison of the Nordic countries shows very little relation between the interval between the screening rounds and reduction of risk, or very little relation between the target age range and reduction of risk. This provides confirmation of the IARC working group [3,19,20] conclusion on the basis of several large-scale programs that the protective effect of screening is high for screening intervals up to five years and for a lower age limit up to 30 years. Organized programs, as contrasted to opportunistic ones, promote adequate quality control [21] and high attendance (e.g., by personal letters of invitation and of response). High coverage and attendance seem to be the single most important determinant of successful screening. Opportunistic screening has problems in catching those who would benefit most from screening. Recently Gustafsson et al. [22] suggest a more optimistic view of the efficiency of opportunistic screening. Their conclusion was based on detection rates of in situ carcinomas in Sweden, not on incidence of invasive disease. Such an analysis will not provide evidence for the effectiveness of opportunistic screening. In Finland it was confirmed by a case–control study at the individual level that effectiveness was better for the organized screening than for the spontaneous smear-taking activity [23].

3. Efficiency

In spite of the coverage of the total target population, screening for cervical cancer can be relatively inexpensive; those programs with the largest effect have been low in cost. It seems that screening starting at the age of 25 or even at the age of 30, repeating the smears at five-year intervals, and having an upper age limit of 60 years will provide practically maximal reduction in the risk of cervical cancer. The program assumes 7 or 8 smears during the woman's lifetime. Such a program compares favorably with those that start at age 20 with annual smears, which result in a total of 40 or more tests during the woman's lifespan.

Screening for cervical cancer is relatively inexpensive compared also to the economic costs of screening for cancers of other primary sites. In fact, the costs of screening will be roughly compensated by savings due to more frequent treatment of early disease compared to the treatment costs of cancers detected through normal practice without screening [24].

4. Equity

Equity is the third dimension, in addition to effect and cost, in health services activities. Often there is a tradeoff between effectiveness, efficiency, and equity. Screening for cervical cancer is an exception. As pointed out above, the

Table 5. Age-adjusted incidence rates (per 100,000 woman years) of invasive cervical cancer among Finnish 35–64-year-old women in 1971–1975 and 1981–1985 by social class (I highest, IV lowest) [26]

	Social class			
Period	I	II	III	IV
1971–1975	11	19	22	28
1981–1985	8	8	9	14

effect in terms of reduction in risk is in practice inversely related to cost: programs with a large reduction in risk are based on relatively few smears. The Finnish program is also an example of an effective program with improvement in equity, measured by the outcome (reduction in risk) in different population groups. In the mid-1980s, the risk was high in remote areas and in lower social classes. Some of the remote areas have benefited most [25], and at the same time the social class differences were reduced. In the early 1970s, the relative risk between the lowest and highest social class was 2.6, as compared to 1.7 in the beginning of the 1980s. The difference in risk between social classes had disappeared, except that the lowest class still was at higher risk than the others (table 5).

5. Conclusion

While the effectiveness of screening for cervical cancer can be relatively reliably evaluated, some of the adverse effects are more problematic. First, some lesions fulfill the histological criteria of malignancy but lack the malignant potential to kill the woman during her lifetime, resulting in overdiagnosis and sometimes overtreatment. The frequency of such lesions depends on the transition probabilities from dysplasia to carcinoma in situ and from carcinoma in situ to invasive disease. Estimating the probabilities by follow-up of small groups of patients will inevitably lead to bias, e.g., because the diagnostic maneuver may destroy the lesion. It is more reliable to estimate the lifetime risks of such lesions by epidemiological means on the basis of screening materials, and to compare the lifetime risks of invasive cancer of those with different preinvasive lesions. More important, however, the transition probabilities are a reflection more of the local diagnostic practices than of any general biological phenomena. In Finland, the diagnostic standard adopted results in CIN III lesions, of which one in three would have progressed into invasive disease if left untreated [6]. In Sweden, the transition probability is substantially smaller, after several years, the annual number of carcinoma in situ lesions was about five times the number of invasive cancers before screening [27].

Second, the quality-of-life effects are poorly known and research is rare. Often the women who attend screening do so to be reassured [28] rather than to reduce the risk of death, i.e., the women attend because of quality-of-life reasons. Important research on anxiety caused by invitation, a false or correct positive test, and other quality-of-life aspects have been and are currently being conducted, especially in the U.K. [29,30].

To screen or not to screen for cervical cancer does not depend on poor information about biological effects or on organizational aspects. The decision depends on balancing the effect on the length of life (which is relatively well known and easy to establish), on the quality of life (which is poorly known and difficult to measure), and on the cost (which again, is relatively easy to measure, but relatively poorly known). It is likely that experts give greater value or weight to the length of life, women value the quality of life, while those responsible for administration are cost conscious. Therefore, the decision to establish and continue screening progams depends not only on the factual evidence available but also on whose values of the benefits, harms, and costs prevail.

References

1. Miller AB, Knight J, Narod S. 1991. The natural history of cancer of the cervix, and the implications for screening policy. In Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.), Cancer screening. UICC Project on Evaluation of Screening for Cancer. Cambridge: International Union Against Cancer, pp. 141–152.
2. Fidler HK, Boyes DA, Worth AJ. 1968. Cervical cancer detection in British Columbia. J Obstet Gynaecol Br Commonw 75:392–404.
3. Day NE. 1986. The epidemiological basis for evaluating different screening policies. In Hakama M, Miller AB, Day NE (eds.), Screening for cancer of the uterine cervix. Lyon: IARC Scientific Publications No. 76, pp. 199–212.
4. Hakama M, Magnus K, Petterson F, Storm H, Tulinius H. 1991. Effect of organized screening on the risk of cervical cancer in the Nordic countries. In Miller A, Chamberlain J, Day N, Hakama M, Prorok P (eds.), Cancer Screening. Cambridge: Cambridge University Press, pp. 153–162.
5. Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC. 1985. Evaluation of screening programmes for gynaecological cancer. Br J Cancer 52:669–673.
6. Hakama M, Rasanen-Virtanen U. 1976. Effect of a mass screening program on the risk of cervical cancer. Am J Epidemiol 103:512–517.
7. Hakama M. 1982. Trends in the incidence of cervical cancer in the Nordic countries. In Magnus K (ed.), Trends in Cancer Incidence, Causes and Practical Implications. New York: Hemisphere, pp. 279–292.
8. Aareleid T, Pukkala E, Thomson H, Hakama M. 1993. Cervical cancer incidence and mortality trends in Finland and Estonia: a screened vs. an unscreened population. Eur J Cancer 29A:745–749.
9. Johannesson GE, Geirsson G, Day N, Tulinius H. 1982. Screening for cancer of the uterine cervix in Iceland 1965–1978. Acta Obstet Gynecol Scand 61 (Suppl):199–203.
10. Petterson F, Bjorkholm E, Naslund I. 1985. Evaluation of screening for cervical cancer in Sweden: trends in incidence and mortality 1958–1980. Int J Epidemiol 14:521–527.
11. Lynge E. 1983. Regional trends in incidence of cervical cancer in Denmark in relation to local smear-taking activity. Int J Epidemiol 12:405–413.

12. Lynge E, Madsen M, Engholm G. 1989. Effect of organized screening on incidence and mortality of cervical cancer in Denmark. Cancer Res 149:2157–2160.
13. Pedersen E, Hoeg K, Kolstad P. 1971. Mass screening for cancer of the uterine cervix in Ostfold county, Norway: an experiment. Second report of the Norwegian Cancer Society. Acta Obstet Gynecol Scand Suppl 11:1–18.
14. Magnus K, Langmark R, Andersen A. 1987. Mass screening for cervical cancer in Ostfold country of Norway 1959–1977. Int J Cancer 39:311–316.
15. Engeland A, Haldorsen T, Tretli S, Hakulinen T, Horte LG, Luostarinen T, Magnus K, Schou G, Sigvaldason H, Storm HH, Tulinius H, Vaittinen P. 1993. Prediction of cancer incidence in the Nordic countries up to the years 2000 and 2010. Acta Pathologica, Microbiologica et Immunologica Scandinavica 101 (Suppl 38).
16. Hakulinen T, Andersen A, Malker B, Pukkala E, Schou G, Tulinius H. 1986. Trends in cancer incidence in the Nordic countries. Acta Path Microbiol Immunol Scand Sect A 94 (Suppl 288):1–151.
17. Laara E, Day N, Hakama M. 1987. Trends in mortality from cervical cancer in the Nordic countries: association with organized programmes. Lancet i:1247–1249.
18. Hristova L, Hakama M. In press. Effect of screening for cancer in the Nordic countries on deaths, costs and quality of life up to 2017. Acta Oncologica, supplement.
19. IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. 1986. Screening for squamous cervical cancer: the duration of low risk after negative result of cervical cytology and its implication for screening policies. Br Med J 293:659–664.
20. IARC Working Group on Cervical Cancer Screening. 1986. Summary chapter. In Hakama M, Miller AB, Day NE (eds.), Screening for Cancer of the Uterine Cervix. Lyon: IARC Scientific Publications No. 76, pp. 133–142.
21. Sigurdsson K. 1995. Quality assurance in cervical cancer screening: The Icelandic experience 1964–1993. Eur J Cancer 31A:728–734.
22. Gustafsson L, Sparen P, Gustafsson M, Wilander E, Bergstrom R, Adami HO. 1995. Efficiency of organized and opportunistic cytological screening for cancer in situ of the cervix. Br J Cancer 72:498–505.
23. Nieminen P, Kallio M, Hakama M. In preparation. Effectiveness of the organized and spontaneous screening for cervical cancer.
24. Hristova L. 1995. Effect of screening for cancer on mortality, costs and quality of life in Finland. Acta Universitatis Tamperensis A 456. Tampere: University of Tampere.
25. Hakama M, Kallio M, Pukkala E. 1995. Kohdunkaulan syovan seulonnat — vaikuttavia tai oikeudenmukaisia? Suom Laakaril 50:2527–2533.
26. Pukkala E. 1995. Cancer risk by social class and occupation. A Survey of 109,000 cancer cases among Finns of working age. In Contributions to Epidemiology and Biostatistics, vol. 7. Basel: Karger.
27. Sveriges Officiella Statistik. 1976. Gynekologisk halsoundersokning 1967–1973. Statistika meddelanden HS, 1.
28. Kauppinen M, Kauraniemi T, Koli T, Voipio N. 1970. Response to the written invitation in a gynaecological mass screening by cytology arranged in Helsinki in 1966. Acta Obstet Gynaecol Scand 49 (Suppl 7):1–20.
29. Campion MJ, Brown JR, McCance DJ, et al. 1988. Psychosexual trauma of an abnormal cervical smear. Br J Obstet Gynaecol 95:175–181.
30. Posner T, Vessey M. 1988. Prevention of cervical cancer. The Patient's View. King Edward's Hospital Fund for London. London: King's Fund Publishing Office.

5. Advances in screening for colorectal cancer

Jack S. Mandel

1. Introduction

There are relatively few established etiologic factors for colorectal cancer, and therefore primary prevention remains speculative. Diet has been most widely studied, and results indicate that diets high in fat, meat, and protein are positively associated with colorectal cancer, whereas diets high in vegetables, fruits, and fibers are negatively associated [1]. Intervention studies have yet to show that modifying diets will alter risk.

Secondary prevention through annual screening for fecal occult blood has been shown to reduce colorectal cancer mortality [2]. This chapter provides an overview of the scientific evidence on screening for colorectal cancer. Particular emphasis is placed on studies of fecal occult blood tests (FOBTs) and flexible sigmoidoscopy.

The development of a fecal occult blood test for home use for colorectal cancer screening was initiated by Greegor in 1967 [3]. Hemoccult® has been the most widely studied FOBT, although persuasive evidence of its effectiveness in reducing colorectal cancer mortality was not available until 1993 [2,4].

2. Screening for blood in stool

2.1. Randomized controlled trials

There are four randomized controlled trials of FOBTs (table 1). Two are completed (Minnesota and Funen), although only one has published results to date. In the Minnesota trial, a total of 46,551 participants aged 50 to 80 were randomly assigned to annual screening, biennial screening, or a control group [2,5]. Participants in the screened groups submitted six guaiac-impregnated paper slides (Hemoccult®), two from each of three consecutive stools, while maintaining a diet free of red meat, poultry, and fish, certain vegetables and fruits, and discontinuing the use of vitamin C tablets and aspirin for at least 24 hours before and during the collection of the samples. Most screening tests were rehydrated with a drop of deionized water to restore the sensitivity,

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

Table 1. Summary of randomized controlled trials

	Minnesota, U.S.A.	Göteborg, Sweden	Nottingham, England	Funen, Denmark
Date started	1975	1982	1981	1985
Number of participants	46,551	68,308	152,928	61,938
Age group	50–80	60–64	50–74	45–74
Screening test	Hemoccult	Hemoccult II[a]	Hemoccult	Hemoccult II
Slides rehydrated	Yes	Yes	No	No
Diet restrictions	Yes	Yes	Yes[b]	Yes
Frequency of screening	Annual, biennial	20–24 months	Biennial	Biennial
Number of screens	Annual (11), biennial (6)	2	5	5
Compliance with first screen (%)	85	63	53	67
Rescreening compliance (%)	75	60	77[c]	93[c]
Slides positive (%)				
Rehydrated	9.8	6	—	—
Not rehydrated	2.4	1.9	2.1	0.9
Sensitivity for colorectal cancer (%)	92[d,e]	83[d,f]	68[g]	48
Specificity for colorectal cancer (%)	90[d]	96[d]	98	99
Positive predictive value for colorectal cancer (%)	2.2[d]	5.2[d]	12.2	8.2

[a] Same guaiac-based test as Hemoccult, except two samples are collected instead of one.

[b] On retesting after a positive test.

[c] Rescreening offered only to those who complied with initial screen.

[d] Rehydrated slides.

[e] False negatives or interval cancers defined as those occurring within one year after a negative screen.

[f] False negatives or interval cancers defined as those occurring 16 to 24 months after a negative screen.

[g] False negatives or interval cancers defined as those occurring within two years after a negative screen. Sensitivity was 65% if tests were collected over three days and 74% if collected over six days.

which was reduced because of drying. Participants with a positive screening test underwent a diagnostic evaluation that included colonoscopy. During the course of the study, 81% of the test positives had a colonoscopy, 12% had other diagnostic tests, 2% repeated the Hemoccult®, and 5% had no follow-up.

Follow-up for vital status through year 13 was 100% complete. The cumulative colorectal cancer (CRC) mortality rate per 1000 population was significantly lower in the annually screened group compared to the control group. The cumulative CRC mortality rate for the biennial group was not significantly decreased, due in part to an increase in the number of deaths in the early years of the study, which was consistent with chance.

The five-year survival rate was highest in the annually screened group and lowest in the control group. The biennial group was intermediate. The screen-detected cases in both screening groups had the highest survival rate; the

survival for the nonscreen-detected cases was considerably lower and similar to the control group.

The stage distribution by study group was consistent with the mortality results. There were twice as many Dukes D cancers in the control group as in the annually screened group. The five-year survival for Dukes D was 2.5%, and these cancers accounted for about one half of the colorectal cancer deaths. Clearly, the marked reduction in Dukes D cancer in the screened group contributed substantially to the lower colorectal cancer mortality rate.

The observed 33% mortality reduction in the annual group underestimated the true reduction, since only 46% of the participants in the annual group completed all of the screens; 10% did not do any screens, and 17% of the tests were not rehydrated. In addition, about 45% of the control group received at least one FOBT in the mid- to late 1980s through their regular physicians. These factors resulted in a lower mortality reduction than potentially achievable. The maximum reduction in fully compliant individuals might exceed 40%.

Because of the high positivity rate (10%), the question was raised as to the independent contribution of colonoscopy in detecting nonbleeding cancers among false positives [2]. During the 13-year period of the study, 32% of those in the annual screened group had a colonoscopy. Many of these colonoscopies were performed in the latter years of the study, when their effect on mortality would have been small. By the ninth year of follow-up, 19% had a colonoscopy, and the mortality reduction was 26%. Lang and Ransohoff [6], using estimated data from the trial, inferred incorrectly that the detection of nonbleeding cancers through random colonoscopy accounted for 35% to 55% of the observed 33% mortality reduction. When actual rather than estimated data were used and an error in their model was corrected, chance detection of nonbleeding cancers ("random colonoscopy") accounted for only 8% to 11% of the 33% mortality reduction. Detection of bleeding cancers using Hemoccult® accounted for the remaining 89% to 92% [7].

A number of other observations support the role of FOBT in accounting for the observed mortality reduction. The positive predictive value (PPV) for colorectal cancer increased with the number of positive Hemoccult® tests out of the set of six. For all polyps, the PPV for one positive test was about the same as for six positive tests; however, for polyps more likely to bleed (>1 cm), the PPV increased with the number of positive tests. These results suggest that the FOBT detected bleeding lesions.

An additional finding in support of the role of FOBT was the significantly fivefold greater cumulative colorectal cancer incidence rate among those with a positive FOBT on the first screen compared to those with a negative FOBT on the first screen or those in the control group.

Three other randomized clinical trials of FOBT are still in process. In Goteborg, Sweden, 68,308 people born between 1918 and 1931 were randomly allocated to a control or a screen group [8–11]. The screen consisted of three Hemoccult II® guaiac-impregnated paper tests prepared by taking two

samples from each of three consecutive stools while maintaining a diet free of peroxidase-rich vegetables, food containing blood, and vitamin C and iron supplements. One rescreen was conducted 20 to 24 months after the initial screen. Half the slides from the first screen and all the slides from the second screen were rehydrated. The examination of test-positive screenees included a digital rectal examination, proctoscopy, 60-cm sigmoidoscopy, and a double-contrast barium enema.

Following the two screens, there were 117 colorectal cancers and 419 adenomas diagnosed among screen group participants and 44 cancers and 51 adenomas among control group participants [11]. Significantly more Dukes A cancers and significantly fewer Dukes D cancers were found in the screen group than in the control group. In the follow-up period after screening ended, fewer cancers and adenomas were diagnosed in the screen group. There were no differences in the Dukes distribution for screen- and control-group cancers diagnosed during the follow-up period. To date, no reduction in CRC mortality has been reported.

In the Nottingham Colorectal Cancer Screening trial, which started in 1984, 152,928 men and women aged 50–74 were enrolled and randomized [12–16]. Compliers with the initial screen (53%) were rescreened every two years with Hemoccult®. Initially, tests were carried out over three consecutive days without dietary restrictions. After 1985, participants with a positive screening test repeated the screening test over a six-day period while excluding red meats and vegetables high in peroxidase from their diet. The diagnostic protocol included a digital rectal examination, 60-cm sigmoidoscopy, and double-contrast barium enema. Later in the study, those with a second positive screening test were colonoscoped. Subjects with a negative second test prepared while on a restricted diet were asked to repeat the test later while maintaining a restricted diet.

The proportion of Dukes A cancers was significantly lower in the control group (14%) than in the screen-detected cases (47%) and in the entire screened group (28%). The proportion of Dukes D cancers was significantly higher in the control group (21%) than in the screen-detected group (6%), but similar in the entire screened group [15]. The trial is still in progress, and mortality data are not yet available.

In Funen, Denmark, 61,938 residents, aged 45 to 75, were randomly assigned to a biennial screen or a control group [17–22]. Six Hemoccult II® slides were obtained from three consecutive stools while participants maintained a restricted diet consisting of no red meat or fresh fruit and abstained from using iron supplements, vitamin C, aspirin, or other nonsteroid antirheumatics during the three-day testing period.

There were five screens at two-year intervals. Only compliers with the initial screen (67%) were invited to participate in a subsequent screen. Those with a positive screening test were colonoscoped. There were 20% fewer colorectal cancer deaths among screen-group participants than among control-group participants. When the deaths were divided among those which oc-

curred in subjects screened and those in nonresponders, the disparity increased. The proportion of CRC deaths among screenees, nonresponders, and control group participants was 0.17%, 0.34%, and 0.30%, respectively.

There was a greater percentage of Dukes A cancers in the screened group than in the control group (25% versus 9%). There was a more favorable stage distribution among screen-detected cancers (51% Dukes A, 6% Dukes D) than among interval-detected cancers (18% Dukes A, 22% Dukes D), cancers in nonresponders (11% Dukes A, 41% Dukes D), and cancers among control-group participants (9% Dukes A, 24% Dukes D). Interval cancers were generally larger than screen-detected cancers, and cancers ≤ 2 cm occurred more often among those that were screen detected.

The number of study participants in the trials ranged from 46,551 to 152,928 (table 1). All the trials used a fecal occult blood test with dietary restrictions. In Nottingham, the dietary restriction was utilized only on retesting of initial test positives. Minnesota and Goteborg rehydrated most of the slides, which resulted in a higher positivity rate, higher sensitivity, lower specificity, and a lower positive predictive value. Compliance varied considerably, from 53% in Nottingham to 85% in Minnesota on the first screen and from 41% in Nottingham to 75% in Minnesota on rescreening when all randomized participants are considered. Screening intervals varied as well. Minnesota screened annually and biennially. The other trials were essentially biennial intervals (Goteborg was 20 to 24 months). There was a wide variation in positivity rates (1.0% to 9.8%), which was due to at least two factors, namely, the age groups tested and rehydration of the tests. The latter was the more significant contributor to the positivity rate.

Minnesota is the only study to date that has published a mortality reduction. Through 13 years of follow-up, there was a statistically significant 33% reduction in colorectal cancer mortality in the annual group compared to the control group [2]. The results from the Danish trial, showing a significant 20% mortality reduction, have been presented but not published [22].

2.2. Nonrandomized controlled studies

There are two prospective controlled but nonrandomized FOBT studies; one in the Burgundy area of France [23,24] and one in New York, U.S.A. [25,26].

The Burgundy study, initiated in 1988, involved 91,000 participants aged 45 to 74 years [23,24]. All residents of certain towns and administrative districts were offered the screening test; residents of the neighboring areas were not offered the screening test.

The first screen was conducted in 1988 and 1989, the second took place in 1990, and the third was scheduled to be conducted two years following the second. Participants completed Hemoccult® tests from three consecutive stools without adhering to any dietary restrictions. Those with a positive screening test were offered colonoscopy. This study is still ongoing, and results are not yet available.

A controlled nonrandomized study conducted between 1975 and 1984 by the Memorial Sloan-Kettering Cancer Center in collaboration with the Preventive Medicine Institute (PMI)-Strang Clinic in New York was designed to evaluate the effectiveness of FOBT as a supplement to 25-cm rigid sigmoidoscopy [25,26].

Between 1975 and 1979, 21,756 clinic patients over the age of 40 were enrolled. There were two groups of patients coming to the clinic; one that responded regularly to reminders for annual checkups (Regulars) and another that came for single visits, often because of specific health concerns (First Timers).

The Regulars and First Timers were separately allocated, based on enrollment date, to study and control groups. The study group prepared three Hemoccult® tests while adhering to a meat-free, high-bulk diet for one day prior to, and for three days during, slide preparation. Participants with a positive FOBT received double-contrast barium enema and full colonoscopy. All study participants underwent the standard PMI-Strang Clinic examination at enrollment, which included a 25-cm rigid proctosigmoidoscopy. Those with a polyp 3 mm in size or larger were referred for colonoscopy.

The prevalence rate of colorectal cancer was higher in the study groups than in control groups, whereas incidence rates were approximately equal. For the prevalent cancers, there was a greater proportion of Dukes A and a smaller proportion of Dukes D in the study groups compared to the control groups. The stage distribution was more similar for incident cancers, which was probably due to the low compliance, which resulted in very few screen-detected cancers.

Survival was significantly better (70% versus 48%) for the First Timers study than the control group. There was no difference between the Regulars study and control groups. There was a 43% reduction ($p = 0.053$) in colorectal cancer mortality between the First Timers study and the control group. This mortality reduction occurred among all age groups, but was somewhat greater among those over 65 than those under 65 years of age.

Four case–control studies on FOBT have also been published [27–30]. Despite the potential problem of selection bias in case–control studies of cancer screening, these studies can nevertheless be considered in evaluating whether the results generally support findings from randomized controlled trials.

A case–control study of FOBT screening conducted at the Kaiser Permanente Medical Care Program (KPMCP) of Northern California included 486 cases who were members age 50 or over, who were diagnosed with colorectal adenocarcinoma between 1981 and 1987, and who subsequently died from the disease before December 1988 [27]. Controls matched to the cases on age, sex, and date of health plan entry were randomly selected from KPMCP membership lists. For 96 cases with distal location of the colorectal cancer, four controls were selected for a separate analysis of screening sigmoidoscopy. An adjusted odds ratio of 0.69 (95% CI, 0.52–0.91) was ob-

served for exposure to at least one screening FOBT during the five-year interval prior to diagnosis.

A population-based case–control study of FOBT was conducted in the Saarland in Germany in men and women aged 45 years or over [28]. The FOBT was introduced in the Federal Republic of Germany in 1977 and was offered to all men and women aged 45 years or over. There were 429 cases, aged 55–74 years, who died between 1983 and 1986 and who were initially diagnosed with colorectal cancer between 1979 and 1985. Up to five age-matched controls were identified from the files of the physician of the corresponding case. Thirteen percent of male cases and 14% of male controls had at least one asymptomatic FOBT 6 to 36 months prior to diagnosis (odds ratio = 0.92; 95% CI, 0.54–1.57). For the same prediagnostic period, 16% of female cases and 29% of female controls had at least one asymptomatic FOBT (odds ratio = 0.43; 95% CI, 0.27–0.68). For the period 12 to 36 months prior to diagnosis, the corresponding odds ratios were 0.73 (95% CI, 0.40–1.32) for men and 0.40 (95% CI, 0.25–0.65) for women. Thus, there was an 8% to 27% nonsignificant benefit for men and a 57% to 60% significant benefit for women.

Newcomb et al. [29] conducted a case–control study using 66 members of the Greater Marshfield Community Health Plan who died of colorectal cancer from 1979 to 1988 and 196 members matched on gender, age, and enrollment duration whose records were reviewed for a history of screening for colorectal cancer. Only screening tests done in the absence of symptoms were considered.

FOBT screening was not associated with a lower colorectal cancer mortality (odds ratio = 1.15; 95% CI, 0.93–1.44). However, only 21% of the cases and 16% of the controls received multiple-slide evaluation of stool. Most screening was done on a single sample obtained during a digital rectal examination.

In the Puget Sound case–control study, cases were members of an HMO who died of colorectal cancer [30]. Matched controls were randomly selected from a list of HMO members. The odds ratio for those ever screened was 0.71 (95% CI, 0.50–1.00) for home use of FOBT and 0.95 (95% CI, 0.67–1.36) for office use of FOBT.

There is now evidence from two randomized controlled clinical trials, a prospective nonrandomized study, and three case-control studies that screening for FOBT with Hemoccult® can reduce colorectal cancer mortality from 20% to 57% (table 2). The Marshfield study is the only published study that did not find a benefit from FOBT. As mentioned earlier, the FOBTs were primarily single tests done as a part of a digit rectal examination.

The results from the two randomized controlled trials are particularly noteworthy, since these type of studies provide the strongest evidence of screening effectiveness. The Minnesota trial found a 33% colorectal cancer mortality reduction, which was diluted by noncompliance. With fully compliant study groups, the mortality reduction would have been greater. To achieve this reduction, the Minnesota trial used mainly rehydrated screening tests with a

Table 2. Results from studies of FOBT

Study	Results in screened group relative to control group		
	Mortality	Survival	Staging
Randomized controlled trials			
Minnesota, U.S.A.	33% reduction[a]	Improved	Favorable shift
Funen, Denmark	20% reduction[a]	Improved	Favorable shift
Nottingham, England	Pending	Improved	Favorable shift
Göteborg, Sweden	Pending	Improved	Favorable shift
Prospective controlled trials			
New York, U.S.A.	43% reduction[b]	Improved	Favorable shift
Burgundy, France	Pending	Pending	Pending
Case–control studies			
Saarland, Germany	57% reduction[a]		
Marshfield, U.S.A.	0% reduction[c]		
California, U.S.A.	31% reduction[a]		
Puget Sound, U.S.A.	29% reduction[d]		

[a] Statistically significant.
[b] $p = 0.053$
[c] 80% of FOBTs done as digital rectal and not as routine screening.
[d] 35% reduction ($p < 0.05$) for those age <75.

positivity rate of 10%, offered 11 screens annually, and averaged 75% compliance with each screen. About 45% of control-group participants had at least one FOBT. In contrast, the Danish trial, which had about a 20% mortality reduction, used nonrehydrated screening tests with a positivity rate of 1%, offered five screens biennially, and averaged less than 60% compliance with each screen. This result is also diluted by noncompliance, although somewhat less so than the Minnesota trial because screening among control-group participants was rare in the Danish trial. From these studies, it appears that the range in the benefit from screening is from about 20% to about 40%, depending on the frequency of screening (biennial versus annual), the number of screens (5 to 11), and the processing of the test (nonrehydrated versus hydrated). It is clear, therefore, that the weight of evidence supports a recommendation for screening with the Hemoccult® test. Although there are preliminary data that FOBTs other than Hemoccult® may be more sensitive and specific, only Hemoccult® has been properly evaluated to date.

2.3. Fecal occult blood tests

The three principal types of fecal occult blood tests are immunochemical, hemeporphyrin, and guaiac (for example, Hemoccult®). Immunochemical or immunological tests are generally qualitative tests, utilizing antibodies directed against the intact globin moiety of human hemoglobin and possibly also against large fragments of globin [31]. They are the most selective of the three types of FOBTs in that they detect only hemoglobin and globin, and

perhaps early degradation products of globin. They are not affected by dietary factors.

A number of studies have evaluated immunochemical tests in selected groups of individuals [32–39]. Although preliminary data are promising, there is insufficient evidence to recommend any of the immunochemical tests for mass screening. Nevertheless, Japan, in 1992, under the auspices of the Health and Medical Services Law for the Aged, implemented colorectal cancer screening using a two-day immunological FOBT [40]. The basis for the recommendation was the result from an unpublished case–control study by Hisamichi that found an odds ratio of 0.36 (95% CI, 0.19–0.68).

Recently a case–control study in Japan by Sato et al. [41] indicated that screening using a one-day immunochemical hemagglutination test (HemeSelect) resulted in a 50% to 60% reduction in colorectal cancer mortality. Descriptive studies evaluating test performance suggest that the immunochemical tests are more sensitive and more specific than other types of FOBTs [32,33,36,38,42,43]. Further study is needed to establish that these tests are superior to the guaiac tests for colorectal cancer screening.

The hemeporphyrin tests detect the broadest range of blood derivatives, namely de-ironed hemes (heme-derived porphyrins) as well as intact heme in any form (free or as hemoprotein). HemoQuant is a biochemical method for the assay of fecal heme and its degradation products in stool [44]. This test showed early promise [44–46]; however, more recent studies have shown that this test is probably less sensitive and less specific than Hemoccult®, and therefore its use for colorectal cancer screening is not recommended [38,47,48].

The performance of any given type of FOBT in a screening program is likely to be dependent on the nature of blood derivatives present in the feces in a given clinical situation [49]. Only the guaiac tests (namely, Hemoccult®) have been sufficiently studied so that enough data are available on which to base a recommendation for screening.

Guaiac tests, which detect heme in any form provided that the iron has not been removed from the porphyrin ring, will react to any peroxidase. Heme, an iron compound of protoporphyrin, constitutes the pigment portion or protein-free part of the hemoglobin molecule. Its fate in the gut is not well understood. However, fecal excretion of heme can be an indicator of gastrointestinal mucosal pathology [50]. The amount excreted will depend on the balance between the amount delivered to the lumen, including that which is present in the diet as myoglobin and hemoglobin, and its loss from the lumen [51].

In the stomach, heme is rapidly released from the binding protein and precipitated. In small intestine it is solubilized, and about 5% to 15% is absorbed [50]. Globin (a protein from hemoglobin) is rapidly digested in stomach and small intestine. The majority of heme presenting to the upper gut is passed to the colon. Hemoglobin, which enters the large intestine from bleeding in the upper gastrointestinal tract, either remains whole and is ex-

creted as intact hemoglobin in the feces or undergoes proteolytic digestion of the globin to become heme [49].

Protoheme, heme coming from the ileum or from hemoglobin, is either excreted in the feces as intact heme or converted by bacteria to a range of heme-derived porphyrins lacking iron [49]. This conversion is a slow and incomplete process, and the amount converted in this way depends on colonic transit rate, site of bleeding, and amount of luminal heme [51]. In addition to the removal of iron, bacterial activity can cause modifications of the vinyl side chains of heme to produce various iron–porphyrin derivatives that are excreted in the feces [49]. As a consequence, feces contain variable proportions of heme and heme-derived porphyrins [51].

From in vivo and in vitro studies, it is possible to predict the nature of hemoglobin derivatives likely to be found in feces, based on the site of bleeding [49]. The more proximal the bleeding, the more likely are heme-derived porphyrins to be the principal products. The more distal the bleeding, the more likely it is that intact hemoglobin will be present in the feces.

Many conditions can result in blood in stool. These include cancer, polyps, peptic ulcers, hemorrhoids, and diverticulitis [49]. Normal individuals lose 0.5 to 2.0 ml/d blood per day [52]. Patients with colorectal cancer or adenomatous polyps generally (but not always) lose more blood, but not all cancers and polyps bleed, nor is the blood evenly distributed in stool [46,53–56]. Furthermore, detectable bleeding tends to be intermittent [57]. Blood loss from colorectal cancer and polyps is greater in right- than left-sided lesions, is higher in patients with Dukes C or D cancers than in patients with Dukes A or B cancers and is higher in patients with larger than smaller adenomas [33,58–60].

Reported sensitivities of Hemoccult® for colorectal cancer varied from 26% to 92% [10,48,54,56,61–67]. Most of the studies have shown sensitivity values of 65% or greater for colorectal cancer [63]. Other guiaic tests have shown similar sensitivities [68]. For rehydrated Hemoccult®, sensitivities were between 83% and 92% [10,58,66]. Sensitivity of Hemoccult® for adenomatous polyps is lower, ranging from 15% to 30% for all polyps, but higher for larger polyps [38,47,63,69]. Jahn et al. [70] found that Hemoccult® was positive in 20% of patients with adenomas 10 to 19 mm in size and 40% in adenomas greater than 20 mm.

The wide range in sensitivity values is not surprising. Hemoccult® is a test for blood in the stool. It is not a direct test for cancer or polyps. Virtually all cancers and some polyps eventually bleed. Bleeding is intermittent, and the blood is not evenly distributed throughout stool. Therefore, a test for blood in stool should identify most cancers and some polyps if it is applied repeatedly with multiple samples to overcome the intermittent bleeding and the unequal distribution of blood in stool. The more samples that are taken, the greater is the likelihood of detecting the bleeding. The evidence reviewed above shows that screening annually — three to six tests each time for a number of years — is likely to lead to detection of the lesion early enough to significantly reduce the probability of death.

The Minnesota study applied the Hemoccult® test annually to an average risk-asymptomatic population [2]. The high sensitivity for colorectal cancer in the study (80% for nonrehydrated Hemoccult® and 92% for rehydrated Hemoccult®) is not surprising, given the frequent application of the test (11 annual screens, 6 tests at each screen) and the population screened (age 50 to 80). Cancer screening tests are generally recommended for repeated (e.g., annual) application. Because the transition from normal mucosa to invasive carcinoma takes many years [71,72], repeat screening increases the probability of detecting a lesion and the probability of detecting it early enough to reduce mortality.

Specificity refers to the ability of the test to identify those without the disease of interest (true negatives). False positives are costly because of the unnecessary diagnostic procedures. The specficity of Hemoccult® for colorectal cancer is 97% to 98% [62,63,66,73]. It is lower (90%) for rehydrated slides [66].

Many foods, including red meats, have peroxidase-like activities that can cause false-positive results on screening tests. Dietary recommendations for fecal occult blood testing vary from no restriction to exclusion of all types of meat or red meat plus uncooked fruit and vegetables [61,62,74]. Overall, the results from the studies of diet indicate that foods can interfere with the test and, therefore, dietary restriction just before and during the testing period should be recommended.

3. Screening for adenomatous polyps

Screening with 60-cm flexible sigmoidoscopy has received considerable attention [75–78]. The rationale for using this procedure relates to the direct visualization and subsequent removal of neoplastic lesions to prevent the development of cancer (in the case of polyp removal) or to prevent the progression of cancer (in the case of cancer removal). The underlying assumption to support this approach is that most, and perhaps all, cancers arise from polyps, and therefore removal of polyps, along with continued surveillance to identify and remove new or recurrent polyps, will prevent cancer. The data in support of this hypothesis are speculative but reasonably persuasive.

3.1. Adenoma–carcinoma sequence

The idea that the majority of colorectal cancers evolve from benign adenomas goes back many years [79–84]. Dukes was among the first to comment, in 1925, on the malignant potential of villous and adenomatous polyps [85]. Even though the evidence is indirect, most investigators believe that the majority of colorectal cancers arise from adenomatous polyps [86–92]. The malignant potential of adenomatous polyps varies with size (cancer is rare in polyps less

than 1 cm) and histology (villous pattern has the most malignant potential) [81,93–98].

It has been hypothesized that colorectal cancer evolves through a sequence of stages beginning with carcinogens acting on a genetically susceptible mucosa and resulting in a hyperproliferative state, followed by a series of oncogene mutations and chromosome deletions [83,99]. This leads to a precursor adenoma, successive stages of dysplasia, and then invasive cancer.

The evidence for the adenoma–carcinoma sequence is indirect, since adenomas are generally resected when discovered rather than left intact and observed to determine their natural history [92]. The present belief that all polyps, at least those greater than 5 mm, should be removed makes it impossible to design a randomized trial to better understand the physiologic process.

The de novo origin of cancer has received more attention in recent years [90]. Some carcinomas appear to have arisen out of flat mucosa and in the absence of adenomatous tissue [100–103]. These lesions are rare [86,87], particularly within the rectum and sigmoid colon. However, there may be two pathways in the development of colorectal cancer, one in which the malignancy develops through benign polypoid growth (the "adenoma–carcinoma sequence") and one in which dysplastic epithelium grows invasively directly from flat mucosa ("de novo") [90].

It takes about 10 years to develop invasive cancer, and at least five years for an adenoma to develop into a cancer [71,72]. The interval is shorter for a single adenoma and longer for multiple adenomas, adenomas with infiltrating cancer, or significant dysplasia. The natural history of 5 to 10 years for the adenoma–carcinoma sequence was suggested by several early studies [81,104,105].

The observations that the prevalence of both adenomas and colorectal cancers increases with age and that colorectal adenoma prevalence at autopsy is correlated with colorectal cancer incidence has lent credence to the adenoma–carcinoma theory [83,86,90–92]. Adenoma prevalence was first assessed in 1947 with a study of 1460 consecutive autopsies in which the entire bowel was examined [106]. Since then, many autopsy surveys have been conducted [107–125].

The frequently repeated demonstration of foci of invasive cancer in adenomas has provided the strongest evidence for the premalignant nature of adenomas [81,90,126]. Residual adenomatous tissue adjacent to a carcinoma provides only circumstantial evidence for the adenoma–carcinoma concept [86]. The natural history of familial polyposis coli also supports the adenoma to carcinoma sequence (see chapter 12, this volume).

The biological changes observed in the development of neoplasia generally reflect alterations in cell proliferation or the regulation of cell proliferation [92]. Many such changes associated with colorectal cancer also have been found in adenomas. Abnormal activation of small ras oncogenes has been demonstrated in both colorectal carcinomas and in colorectal adenomas [127].

The adenoma–carcinoma sequence is also supported by genetic studies that have shown in vitro transformation of benign cell lines to malignant cells. Fearon and Vogelstein's [128] hypothesis of colorectal carcinogenesis postulated a sequence of genetic alterations involving the mutational activation of an oncogene coupled with the loss of several genes that normally suppress tumorigenesis. Initially, there is a chromosome 5 and ras oncogene mutation, then chromosome 18 loss resulting in severe dysplasia in adenomas, and finally chromosome 17 loss before progression to invasive carcinoma.

Genetic studies have shown that cancer cells arising within an adenoma exhibited the identical molecular alterations as the adenoma cells, but in addition acquired mutations that were presumably critical for the malignant phenotype [129]. This latter point is one of the stronger arguments that colon carcinomas derive from preexisting adenomas [91].

Tierney et al. [88] concluded that even though colorectal adenomas lack the capacity for invasion or the metastatic characteristics of a malignant lesion, they shared, to various degrees, the anaplastic characteristics of carcinoma. The loss of cellular control mechanisms in both adenomas and carcinomas is reflected by abnormalities in DNA ploidy, enzymatic activities, expression of carcinoembryonic antigen, and composition of mucin. Neugut et al. [92] concluded there was some evidence for the following abnormalities in both adenoma and carcinomas: DNA aneuploidy, cell kinetics, enzyme activity and expression, carcinoembryonic antigen expression, and presence of abnormal blood group antigens on cancer cell surfaces.

Further support for the adenoma–carcinoma sequence is derived from familial studies that have shown that adenoma patients were more likely to have a family history of colon cancer than controls [130] and that cancer patients were more likely than controls to have a family history of adenomas [131–133]. Data from colonoscopy screening programs of individuals with a family history of colorectal cancer showed that the prevalence of adenomas varied from 12% to 27% [129,131,134–140]. Age-specific prevalence rates of adenomas in high-risk individuals were up to twice the rates in average-risk individuals [140].

Incidence data on adenomas are scarce. A recent study of all patients undergoing colonoscopy in three New York practices found that the cumulative incidence rate of adenomas at 36 months was 16% [141]. In this study, an incident case was defined as a patient with no abnormality on the index colonoscopy and an adenoma on the follow-up colonoscopy. The study patients were not, however, average-risk patients, since most had some indication to warrant colonoscopy. Furthermore, only 20% of the patients with an initial normal colonoscopy underwent a follow-up colonoscopy. Most did so because of persistent, recurrent, or new symptoms.

Cannon-Albright et al. [142] found a distal adenoma incidence rate of 12% in spouses of individuals who were screened because of their family history of colorectal cancer. Eighty-five percent of the spouses, aged 30 to 81 with no history of colorectal cancer or polyps or inflammatory bowel disease, were

examined using 60-cm flexible sigmoidoscopy. The incidence was somewhat higher in men (14%) than women (10%).

Adenomas are considerably more common than carcinomas. Only a small proportion (about 10%) of adenomas give rise to carcinomas [90,143]. The larger adenomas (≥1 cm) are more likely to evolve into carcinomas [81]. Recurrence rates of adenomas range from 15% to 60% [92,141,144–148]. The most recent data indicate a three-year recurrence rate of about 30%. Removal of polyps and ongoing colonoscopy surveillance appear to reduce the incidence of colorectal cancer [64,71,72,98,149].

The National Polyp Study (NPS), a multicenter prospective randomized trial, was designed to evaluate follow-up surveillance strategies in patients who have undergone polypectomy for the control of large bowel cancer [64,71,72]. After colonoscopy, patients were excluded if they had no polyps, non-adenomatous polyps, malignant polyps, a sessile adenoma larger than 3 cm in diameter, or colorectal cancer. Of the 9112 referred patients, 6480 became ineligible at this stage. The remaining 2632 patients had one or more adenomas.

A total of 1418 patients with at least one histologically documented colorectal adenoma who had undergone a complete colonoscopy during which all identified polyps were removed were randomly assigned to more frequent follow-up (1 and 3 years after initial polypectomy) or less frequent follow-up (3 years) with colonoscopy and barium enema.

The expected number of colorectal cancers in the study cohort, based on three difference reference groups (Mayo Clinic, St. Mark's, and SEER rates) were 48.3, 43.4, and 20.7, respectively. The observed number ($n = 5$) of incident colorectal cancer in the study cohort was significantly lower ($p < 0.001$) than the expected number. The standardized incidence ratio was 0.10 (95% CI, 0.03–0.24) for the Mayo Clinic group, 0.12 (95% CI, 0.04–0.27) for the St. Mark's group, and 0.24 (95% CI, 0.08–0.56) for the SEER group.

The authors concluded that the significantly reduced incidence of colorectal cancer provided evidence of the progression of adenoma–carcinoma sequence and of the effectiveness of the current practice of removing adenomatous polyps in the colon. The follow-up period in this study was relatively short, given the natural history of the development of colorectal cancer. Additional follow-up is needed before a definitive conclusion can be reached.

The Funen Adenoma Follow-up Study was designed to reduce the incidence and mortality of colorectal carcinoma in adenoma patients [149]. A total of 1042 patients with colorectal adenomas were allocated to different colonoscopy follow-up intervals ranging from 6 to 48 months. There were 1689 colorectal adenomas removed in 1042 patients. One hundred and eighteen patients had severe dysplasia in one or more adenomas. Of 1689 adenomas, 122 showed severe dysplasia. Size, structure, and anatomic location were independent risk factors for severe dysplasia. The adjusted odds ratio was 9.3 for adenomas of 10–19 mm and 25.2 for adenomas of 20 mm or more relative to small adenomas. Adjusted odds ratios were 1.9 (95% CI, 1.2–3.0) for tubulovillous adenomas and 2.3 (95% CI, 0.7–7.3) for villous adenomas rela-

tive to tubular adenomas. The adjusted odds ratios for anatomic location were 0.7 (95% CI, 0.4–1.0) for the sigmoid colon, 0.4 (95% CI, 0.1–1.6) for the descending colon, and 0.2 (95% CI, 0.1–0.7) for the right colon relative to the rectum.

The St. Mark's Hospital study evaluated the subsequent risk of colorectal cancer following removal of rectosigmoid adenomas [98]. Over 2000 symptomatic patients who underwent excision of one or more rectosigmoid adenomas were followed for up to 30 years to determine their incidence of colorectal cancer. The incidence of rectal cancer was similar to that in the general population (standardized incidence rate = 1.2; 95% CI, 0.7–2.1); however, those with a rectosigmoid adenoma that was tubulovillous, villous, or over 1 cm had 3.6 times (95% CI, 2.4–5.0) the risk of developing colon cancer. Those with multiple rectosigmoid adenomas had 6.6 times (95% CI, 3.3–11.8) the risk of developing colon cancer. Those with only small tubular adenomas (43% of the case group) had no increased risk of colorectal cancer. The authors postulated that a one-time flexible sigmoidoscopy followed by colonoscopic surveillance of those with an adenoma would prevent 550 colorectal cancer cases and 3500 deaths each year in the U.K. However, this estimate was dependent on achieving a compliance rate of 65%.

Removal of adenomas and subsequent colonoscopic surveillance should reduce the incidence and mortality of colorectal cancers. However, FOBT is not a sensitive enough test for adenomas because most adenomas do not bleed.

3.2. Flexible sigmoidoscopy

Flexible sigmoidoscopy (60 cm) has the obvious advantage of direct visualization of adenomas but the disadvantage of examining only the distal portion of the colon. There has not been a randomized controlled clinical trial to demonstrate the effectiveness of this procedure as a screening procedure, yet it is recommended based on mathematical modeling or observational (mainly case–control) studies [76,150], despite the biases with these types of studies [151–153]. Atkin et al. [154] suggest that a one-time flexible sigmoidoscopy at about age 55 could reduce colorectal cancer mortality by about 45%. They propose a randomized trial to evaluate their hypothesis. Such a trial is necessary to provide definitive evidence of the effectiveness of flexible sigmoidoscopy screening. The current Prostate, Lung, Ovarian, and Colorectal Screening Trial, a randomized controlled screening trial in men and women age 55 to 74, should provide definitive evidence as to whether screening an asymptomatic average risk population with 60-cm flexible sigmoidoscopy can significantly reduce colorectal cancer mortality. Results from this trial will not be available for at least eight years.

A few observational studies provide data on the potential benefit of sigmoidoscopy screening. Gilbertsen [155,156] reported on a long-term follow-up of rigid proctosigmoidoscopies of 21,140 adenomatous polyp patients at-

tending a cancer detection center. No rectal cancer deaths were found; however, the study was biased, since the initial rectal cancers were inappropriately excluded [157].

In the Kaiser-Permanente Multiphasic Checkup Evaluation Study [158], 5156 persons aged 35 to 54 years were offered an annual comprehensive screening examination that included a 25-cm rigid sigmoidoscopy for those age 40 and older. A group of 5557 persons assigned to the control group were not encouraged to take any of the tests but were free to do so as part of the Kaiser program. Deaths among the study and control members were ascertained through 1982 by follow-up through the Kaiser records and by matching names of subjects who left the plan against California death records.

Early in the study, there was a lower colorectal cancer mortality rate in the study than control group [159]. By the sixteenth year (1980), there were 12 and 29 colorectal cancer deaths in the study and control groups, respectively, and a stage shift toward earlier cancers in the study group [160]. More sigmoidoscopies were performed on the study group than on control-group members (8.1% vs. 5.2%); however, there was no difference in the frequency with which polyps were removed. The authors concluded that if sigmoidoscopy screening accounted for the mortality reduction, it was due to early detection of cancer rather than to the removal of premalignant adenomas.

Selby et al. [161], in reexamining 8891 medical charts to reconstruct relevant histories that had been erased from computer tapes, concluded that sigmoidoscopy did not account for the reduction in the study-group colorectal cancer incidence because there was no group difference in the rate of detection or removal of polyps; the incidence of colorectal cancer was similar to that expected based on general population data; the mortality difference could be related to small differences in screen-detected cancers, which could be explained by chance, lead time, or length bias; most tumors in both study and control groups were detected because of symptoms, rather than by screening of asymptomatic individuals; the slight improvement in the stage distribution in the study group was not related to screening sigmoidoscopy; and the absence of a substantial difference in sigmoidoscopy between the two groups. The apparent benefit to the study group may have been due to a chance difference between the groups in the prevalence of risk factors for colorectal cancer and the probability of the development of colorectal cancer and thus death from colorectal cancer in the two groups [157].

From two case–control studies it was concluded that a single screening sigmoidoscopy could reduce mortality from cancers of the rectum and distal colon by between 59% and 79% [29,162].

At Kaiser Permanente Medical Care Program (KPMCP) of northern California, cases identified from the Surveillance, Epidemiology and End Results Registry (SEER) were plan members 45 years of age or older diagnosed with colorectal cancer between 1971 and 1987, and who died of the cancer by the end of 1988 [162]. Deaths were ascertained from registry information or by linking the case file to the California death records. This system of case

ascertainment would miss eligible cases who left the area before their cancer was diagnosed, deaths among ascertained cases that occurred outside of California, and deceased cases who did not match the California vital statistics records because of an error in matching variables. It has been estimated that the Kaiser system of matching cases to vital records missed between 8% and 18% of deaths [160].

There were 261 deaths that could have been detected by rigid sigmoidoscopy, i.e., cancers of the rectum or rectosigmoid and cancers of the sigmoid colon that were visualized by rigid sigmoidoscopy or described as within 20 cm of the anus in pathological or surgical reports. A random sample of 268 fatal cancers that were above 20 cm was selected from the remaining cases for separate analysis.

Matched (age, sex, date of entry into the health plan) controls were selected from the health plan membership lists. Controls had to be alive and a member of the health plan when the matched cases died. However, it was not apparent if cases had to be members of the health plan until the time of their death. If not, a bias may have been introduced.

Significantly fewer cases than controls had one or more screening rigid sigmoidoscopies during the 10-year period immediately before the onset of symptoms or the screening test (adjusted odds ratio = 0.41; 95% CI, 0.25–0.69). There was no evidence of screening efficacy for the 268 cases with cancer beyond the reach of the rigid sigmoidoscopy (adjusted odds ratio = 0.96; 95% CI, 0.61–1.50).

This study showed that sigmoidoscopy screening could reduce colorectal cancer mortality by about 30%, assuming that 50% of the colorectal cancers and adenomatous polyps arise within reach of the flexible sigmoidoscope.

Newcomb et al. [29] conducted a case–control study of 66 cases who were members of the Greater Marshfield Community Health Plan (GMCHP) who died of colorectal cancer from 1979 to 1988 and were enrolled in the health plan at least 12 months prior to diagnosis. Controls were 196 randomly selected GMCHP members of the same gender, similar age, and enrollment duration.

The risk of death from colorectal cancer was reduced among individuals having had a single screening sigmoidoscopy (odds ratio = 0.21; 95% CI, 0.08–0.52). The reduction in cancer mortality risk appeared to be limited to cancer of the rectum and distal colon.

The authors noted a number of limitations to their study, including reliance on medial records for screening histories, limited ability to identify potential confounders, and reliance on judgment as to whether symptoms plausibly related to the cancer were present at the time the tests were performed.

Flexible sigmoidoscopy screening has considerable appeal, but there is insufficient evidence to recommend it for mass screening. The observational studies suggest that sigmoidoscopy screening may reduce colorectal cancer mortality; however, these studies are not adequate proof of effectiveness. The two case–control studies could well have overstated the benefit because of

selection bias. For example, in the Selby [162] study, the odds ratio increased from 0.30 to 0.41 by adjusting for three known confounders.

Additional adjustment for other potential confounders could further increase the odds ratio, thus reducing the apparent screening benefit. Screening with flexible sigmoidoscopy can only effect distal and synchronous lesions. It is estimated that only about 50% to 60% of adenomas and cancers would be detected using flexible sigmoidoscopy (108–112,114,118,119,121,163–165]. Furthermore, the percentage of lesions that are proximal and hence beyond the reach of the flexible sigmoidoscope increases with age [112,119,120,166].

Further research is needed, preferably a randomized controlled trial to obtain definitive evidence of the potential benefit of screening with flexible sigmoidoscopy. Consideration must also be given to compliance and cost, particularly the costs and risks incurred from colonoscopy surveillance of patients with adenomatous polyps.

3.3. Colonoscopy

Colonoscopy has been suggested as a one-time screen during the sixth decade of life [167]. Rogge et al. [168] suggested that screening average-risk persons with colonoscopy should start at age 40 and be repeated every five years. The appeal is apparent—visualizing the entire colon could lead to the identification of 95% of the patients with larger adenomatous polyps or cancers. Follow-up examinations would depend on the findings from the baseline colonoscopy. Lieberman [167] proposed no follow-up if no polyps or hyperplastic polyps are found, 10-year follow-up for tubular adenomas, a one-year then five-year follow-up for adenomas greater than 1 cm or villous adenomas, and a one-year, then 3- to 5-year follow-up for carcinomas. His theoretical cost-effectiveness analysis suggested that colonoscopy screening was more cost effective than flexible sigmoidoscopy or FOBT screening.

This idea warrants further exploration. Issues to consider, in addition to cost and effectiveness, are screening frequency, compliance, and risks. There are no data on colonoscopy compliance of an asymptomatic, average-risk population. Additional studies are needed to provide data on effectiveness and compliance before considering colonoscopy as a screening procedure.

4. Recommendations for practice and future research

The only test that has proven to be effective in reducing colorectal cancer mortality is Hemoccult® applied annually to a population over the age of 50. Other fecal occult blood tests show promise, but there is insufficient evidence to recommend their use.

Problems with the Hemoccult® include low sensitivity for adenomas, a large number of false positives, and therefore relatively high cost. If the ongoing randomized trials demonstrate a significant benefit from biennial screening, as

has been demonstrated for annual screening, then the screening cost will be substantially reduced by screening half as often.

Observational studies show a benefit from flexible sigmoidoscopy screening. If these studies are correct, then colorectal cancer mortality could be reduced by about 30%. However, a recommendation for flexible sigmoidoscopy screening is premature because there has not been a definitive study of flexible sigmoidoscopy screening, observational studies may overestimate the benefit, compliance is unknown but likely to be lower than with FOBT, and cost-effectiveness cannot be determined since effectiveness is not known.

Molecular biology may lead to the development of screening tests for the majority of colorectal cancer cases. Recent studies have successfully identified genes in hereditary nonpolyposis colorectal cancer (HNPCC) and familial adenomatous polyposis (FAP) (chapter 12, this volume). These conditions account for about 10% of the colorectal cancer cases. Future work may develop simple cost-effective markers that could be used to identify high-risk persons for colonoscopy [169–171].

References

1. Potter JD, Slattery ML, Bostick RM, Gapstur SM. 1993. Colon cancer: A review of the literature. Epidemiol Rev 15:499–545.
2. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, Ederer F. 1993. Reducing mortality from colorectal cancer by screening for fecal occult blood. N Engl J Med 328:1365–1371.
3. Greegor DH. 1967. Diagnosis of large-bowel cancer in the asymptomatic patient. JAMA 201:943–945.
4. Winawer SJ. 1993. Colorectal cancer screening comes of age. N Engl J Med 328:1416–1417.
5. Gilbertson VA, Church TR, Greive FJ, Mandel JS, McHugh RB, Schuman LM, Williams SE. 1980. The design of a study to assess occult blood screening for colon cancer. J Chron Dis 33:107–114.
6. Lang CA, Ransohoff D. 1994. Fecal occult blood screening for colorectal cancer. JAMA 271:1011–1013.
7. Ederer F, Church TR, Mandel JS. Submitted. Fecal occult blood screening: does chance detection of nonbleeding cancers really play an important role in preventing death from colorectal cancer?
8. Kewenter J, Björk S, Haglind E, Smith L, Svanik J, Ahren C. 1988. Screening and rescreening for colorectal cancer. Cancer 62:645–651.
9. Kewenter J, Asztely M, Engaras B, Haglund E, Smith L, Svanvik J, Ahern C. 1991. A randomized trial of faecal occult blood testing for early detection of colorectal cancer: results of screening and rescreening of 51,325 subjects. In Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.), Cancer Screening. Cambridge, England: Cambridge University Press, pp. 117–125.
10. Kewenter J, Brevinge H, Engaras B, Haglind E, Ahren C. 1994. Follow-up after screening for colorectal neoplasms with fecal occult blood testing in a controlled trial. Dis Colon Rectum 37:115–119.
11. Kewenter J, Brevinge H, Engaras B, Hagind E, Ahren C. 1994. Results of screening, rescreening and followup in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing. Results for 68,308 subjects. Scand J Gastroenterol 29:468–473.

12. Hardcastle JD, Ferrands PA, Balfour TW, Chamberlain J, Amar SS, Sheldon MG. 1983. Controlled trial of faecal occult blood testing in the detection of colorectal cancer. Lancet ii:1–4.
13. Hardcastle JD, Armitage NC, Chamberlain J, Amar SS, James PD, Balfour TW. 1986. Fecal occult blood screening for colorectal cancer in the general population. Results of a controlled trial. Cancer 58:397–403.
14. Hardcastle JD, Thomas WM, Chamberlain J, Pye G, Sheffield J, James PD, Balfour TW, Amar SS, Armitage NC, Moss SM. 1989. Randomised, controlled trial of faecal occult blood screening for colorectal cancer: results for first 107,349 subjects. Lancet i:1160–1164.
15. Hardcastle JD. 1991. Randomized control trial of faecal occult blood screening for colorectal cancer: results for the first 144,103 patients. Eur J Cancer Prev 1:21.
16. Thomas WM, Hardcastle JD. 1991. An update on the Nottingham trial of fecal occult blood screening for colorectal cancer. In Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.), Cancer Screening. Cambridge, England: Cambridge University Press, pp. 106–115.
17. Klaaborg K, Madsen MS, Sondergaard O, Kronborg O. 1986. Participation in mass screening for colorectal cancer with fecal occult blood test. Scand J Gastroenterol 21:1180–1184.
18. Krönborg O, Fenger C, Sondergaard O, Pederson KM, Olsen J. 1987. Initial mass screening for colorectal cancer with fecal occult blood test: a prospective randomized study at Funen in Denmark. Scand J Gastroenterol 22:677–686.
19. Krönborg O, Fenger C, Olsen J, Bech K, Sondergaard O. 1989. Repeated screening for colorectal cancer with fecal occult blood test: a prospective randomized study at Funen, Denmark. Scand J Gastroenterol 24:599–606.
20. Krönborg O, Fenger C, Worm J, Pedersen SA, Hem J, Bertelson K, Olsen J. 1992. Causes of death during the first 5 years of a randomized trial of mass screening for colorectal cancer with fecal occult blood test. Scand J Gastroenterol 27:47–52.
21. Jensen BM, Krönborg O, Fenger C. 1992. Interval cancers in screening with fecal occult blood test for colorectal cancer. Scand J Gastroenterol 27:779–782.
22. Krönborg O, Fenger C. 1994. A randomized population trial with Hemoccult II for colorectal cancer (abstract). World Congress for Gastroenterology, Los Angeles.
23. Faivre J, Arveaux P, Milan C, Durand G, Lamour J, Bedenne L. 1991. Participation in mass screening for colorectal cancer: results of screening and rescreening from the Burgundy study. Eur J Cancer Prev 1:49–55.
24. Arveux P, Durand G, Milan C, Bedenne L, Levy D, Doan BOH, Faivre J. 1992. Views of a general population on mass screening of colorectal cancer: the Burgundy study. Prev Med 21:574–581.
25. Winawer SJ, Schottenfeld D, Flehinger BJ. 1991. Colorectal cancer screening. J Natl Cancer Inst 83:243–253.
26. Winawer SJ, Flehinger BJ, Schottenfeld D, Miller DG. 1993. Screening for colorectal cancer with fecal occult blood testing and sigmoidoscopy. J Natl Cancer Inst 85:1311–1318.
27. Selby JV, Friedman GD, Quesenberry CP, Weiss NS. 1993. Effects of fecal occult blood testing on mortality from colorectal cancer. A case–control study. Ann Intern Med 118:1–6.
28. Wahrendorf J, Robra BP, Wiebelt H, Oberhausen R, Weiland M, Dhom G. 1993. Effectiveness of colorectal cancer screening: results from a population-based case-control evaluation in Saarland, Germany. Eur J Cancer Prev 2:221–227.
29. Newcomb PA, Norfleet RG, Storer BE, Surawicz TS, Marcus PM. 1992. Screening sigmoidoscopy and colorectal cancer mortality. J Natl Cancer Inst 84:1572–1575.
30. Lazovich D, Weiss NS, Stevens NG, White E, McKnight B, Wagner EH. 1995. A case–control study to evaluate efficacy of screening for fecal occult blood. J Med Screening 2:84–89.
31. St. John DJB. 1990. Faecal occult blood tests: a critical review. In Hardcastle JD (ed.), Screening for Colorectal Cancer. Englewood, NJ: Normed Verlag, pp. 55–68.
32. Songster CL, Barrows GH, Jarrett DD. 1980. Immunochemical detection of fecal occult blood. The fecal smear punch-disc test: a new non-invasive screening test for colorectal cancer. Cancer 45:1099–1102.

33. Turunen MJ, Liewendahl K, Partanen P, Adlercreutz H. 1984. Immunological detection of faecal occult blood in colorectal cancer. Br J Cancer 49:141–148.
34. Armitage N, Hardcastle JD, Amar SS, Balfour TW, Haynes J, James PD. 1985. A comparison of an immunological faecal occult blood test Fecatwin sensitive/FECA EIA with Haemoccult in population screening for colorectal cancer. Br J Cancer 51:799–804.
35. St. John DBJ, Young GP, Alexeyeff FM, Deacon M, Cuthbertson AM. 1990. Most large and medium colorectal adenomas can be detected by immuno-chemical occult blood tests. Gastroenterology 98:A312.
36. Castiglione G, Grazzini G, Ciatto S. 1992. Guaiac and immunochemical tests for faecal occult blood in colorectal cancer screening. Br J Cancer 65:942–944.
37. Thomas WM, Hardcastle JD, Jackson J, Pye G. 1992. Chemical and immunological testing for fecal occult blood: a comparison of two tests in symptomatic patients. Br J Cancer 65:618–620.
38. St. John DJB, Young GP, Alexeyeff MA, Deacon MC, Cuthbertson AM, Macrae FA, Penfold JCB. 1993. Evaluation of new occult blood tests for detection of colorectal neoplasia. Gastroenterology 104:1661–1668.
39. Robinson MHE, Marks CG, Farrands PA, Thomas WM, Hardcastle JD. 1994. Population screening for colorectal cancer: a comparison between guaiac and immunological fecal occult blood tests. Br J Surg 81:448–451.
40. Oshima A. 1994. A critical review of cancer screening programs in Japan. Int J Technol Assess Health Care 10:3, 346–358.
41. Saito H, Soma Y, Koeda J, Wada T, Kawaguchi H, Sobue T, Aisawa T, Yoshida Y. 1995. Reduction in risk of mortality from colorectal cancer by fecal occult blood screening with immunochemical hemagglutination test. A case-control study. Int J Cancer 61:465–469.
42. Petrelli N, Michalek AM, Freedman A, Baroni M, Mink I, Rodriguez-Bigas M. 1994. Immunochemical versus guaiac occult blood stool tests: results of a community-based screening program. Surg Oncol 3:27–36.
43. Castiglione G, Sala P, Ciatto S, Grazzini G, Mazzotta A, Rossetti C, Spinelli P, Bertario L. 1994. Comparative analysis of results of guaiac and immunochemical tests for faecal occult blood in colorectal cancer screening in two oncological institutions. Eur J Cancer Prev 3:399–405.
44. Schwartz S, Dahl J, Ellefson M, Ahlquist D. 1983. The "HemoQuant" test: A specific and quantitative determination of heme (hemoglobin) in feces and other materials. Clin Chem 29:2061–2067.
45. Schwartz S, Ellefson M. 1983. Fecal recovery of hemoproteins from blood, meat, and fish ingested by normal volunteers: the HemoQuant assay. Gastroenterology 84:1302.
46. Ahlquist DA, McGill DB, Schwartz S, Taylor WF, Owen RA. 1985. Fecal blood levels in health and disease. A study using HemoQuant. N Engl J Med 312:1422–1428.
47. St. John DJB, Young GP, McHutchison JG, Deacon MC, Alexeyeff MA. 1992. Comparison of the specificity and sensitivity of Hemoccult and HemoQuant in screening for colorectal neoplasia. Ann Intern Med 117:376–382.
48. Ahlquist DA, Wieand HS, Moertel CG, McGill DB, Loprinzi CL, O'Connell MJ, Mailliard JA, Gerstner JB, Pandya K, Ellefson RD. 1993. Accuracy of fecal occult blood screening for colorectal neoplasia. A prospective study of Hemoccult and HemoQuant tests. JAMA 269:1262–1267.
49. Young GP, St. John DJB. 1991. Selecting an occult blood test for use as a screening tool for large bowel cancer. In Rozen P, Reich CB, Winawer SJ (eds.), Large Bowel Cancer: Policy, Prevention, Research and Treatment. Frontiers of Gastrointestinal Research, vol. 18. Basel: Karger, pp. 135–156.
50. Young GP, Rose IS, St. John DJB. 1989. Haem in the gut. I. Fate of haemoproteins and the absorption of haem. J Gastroenterol Hepatol 4:537–545.
51. Young GP, St. John DJB, Rose IS, Blake D. 1990. Haem in the gut. Part II. Faecal excretion of haem and haem-derived porphyrins and their detection. J Gastroenterol Hepatol 5:194–203.

52. Herzog P, Holtermuller KH, Preiss J, Fischer J, Ewe K, Schreiber HJ, Berres M. 1982. Fecal blood loss in patients with colonic polyps: a comparison of measurements with chromium-labeled erythrocytes and with the Haemoccult test. Gastroenterology 83:957–962.
53. Griffith CDM, Turner DJ, Saunders JH. 1981. False negative results of Hemoccult tests in colorectal cancer. Br Med J 283:472.
54. Crowley ML, Freeman LD, Mottet MD, et al. 1983. Sensitivity of guaiac-impregnated cards for the detection of colorectal neoplasia. J Clin Gastroenterol 5:127–130.
55. Rosenfield RE, Kochwa S, Kaczera Z, Maimon J. 1979. Nonuniform distribution of occult blood in feces. Am J Clin Pathol 71:204–209.
56. Doran J, Hardcastle JD. 1982. Bleeding pattern in colorectal cancer: the effect of aspirin and the implications for faecal occult blood testing. Br J Surg 69:711–713.
57. Farrands PA, Hardcastle JD. 1983. Accuracy of occult blood tests over a six-day period. Clin Oncol 9:217–225.
58. Macrae FA, St. John DJB. 1982. Relationship between patterns of bleeding and Hemoccult sensitivity in patients with colorectal cancers or adenomas. Gastroenterology 82:891–898.
59. Ahlquist DA, McGill DB, Fleming JL, Schwartz S, Wieand HS, Rubin J, Moertel CG. 1989. Patterns of occult bleeding in asymptomatic colorectal cancer. Cancer 63:1826–1830.
60. Ahlquist DA, Klee GG, McGill DB, Ellefson RD. 1990. Colorectal cancer detection in the practice setting. Impact of fecal occult blood testing. Arch Intern Med 150:1041–1045.
61. Macrae FA, St. John DJB, Caligiore P, Taylor LS, Legge JW. 1982. Optimal dietary conditions for Hemoccult testing. Gastroenterology 82:899–903.
62. Simon JB. 1985. Occult blood screening for colorectal carcinoma: a critical review. Gastroenterology 88:820–837.
63. Simon JB. 1987. The pros and cons of fecal occult blood testing for colorectal neoplasms. Cancer Metastasis Rev 6:397–411.
64. Winawer SJ, Zauber A, Diaz B. 1987. The National Polyp Study: temporal sequence of evolving colorectal cancer from the normal colon. Gastrointest Endoscop 33:167.
65. Bertario L, Spinelli P, Gennari L, Sala P, Pizzetti P, Severini A, Cozzi G, Bellomi M, Berrino F. 1988. Sensitivity of Hemoccult test for large bowel cancer in high-risk subjects. Dig Dis Sci 33:609–613.
66. Mandel JS, Bond JH, Bradley M, Snover DC, Church TR, Williams S, Watt G, Schuman LM, Ederer F, Gilbertsen V. 1989. Sensitivity, specificity and positive predictivity of the Hemoccult test in screening for colorectal cancers. Gastroenterology 97:597–600.
67. Castiglione G, Grazzini G, Poli G, Bonardi R, Ciatto S. 1991. Hemoccult sensitivity estimate in a screening program for colorectal cancer in the Province of Florence. Tumori 77:243–245.
68. Moran A, Husband D, Jones AF, Asquith P. 1995. Diagnostic value of a guaiac occult blood test and fecal alpha 1-antitrypsin. Gut 36:87–89.
69. Demers RY, Stawick LE, Demers P. 1985. Relative sensitivity of the fecal occult blood test and flexible sigmoidoscopy in detecting polyps. Prev Med 14:55–62.
70. Jahn H, Jorgensen OD, Kronborg O, Fenger C. 1992. Can Hemoccult II replace colonoscopy in surveillance after radical surgery for colorectal cancer and after polypectomy? Dis Colon Rectum 35:253–256.
71. Winawer SJ, Zauber AG, O'Brien MJ, Gottlieb LS, Sternberg SS, Stewart ET, Bond JH, Schapiro M, Panish JF, Waye JD, Kurtz RC, Shike M, Ho MN, and The National Polyp Study Workgroup. 1992. The National Polyp Study. Design, methods, and characteristics of patients with newly diagnosed polyps. Cancer 70:1236–1245.
72. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, Waye JD, Schapiro M, Bond JH, Panish JF, Ackroyd F, Shike M, Kurtz RC, Hornsby-Lewis L, Gerdes H, Stewart ET, and The National Polyp Study workgroup. 1993. Prevention of colorectal cancer by colonoscopic polypectomy. N Engl J Med 329:1977–1981.
73. Windeler J, Kobberling J. 1987. Colorectal carcinoma and Haemoccult. A study of its value in mass screening using meta-analysis. Int J Colorect Dis 2:223–228.
74. Illingworth DG. 1965. Influence of diet on occult blood tests. Gut 6:595–598.

75. Shapiro S. 1992. Case–control studies of colorectal cancer mortality: is the case made for screening sigmoidoscopy? J Natl Cancer Inst 84:1546–1547.
76. Ransohoff DF, Lang CA. 1993. Sigmoidoscopy screening in the 1990's. JAMA 269:1278–1281.
77. Lieberman D. 1994. Colon cancer screening: beyond efficacy. Gastroenterology 106:803–812.
78. Lieberman D. 1994. Screening/early detection model for colorectal cancer. Why screen? Cancer 74:2023–2027.
79. Morson BC. 1974. The polyp–cancer sequence in the large bowel. Proc R Soc Med 67:451–457.
80. Jackman RJ, Mayo CW. 1951. The adenoma–carcinoma sequence in cancer of the colon. Surg Gynecol Obstet 93:327–330.
81. Muto T, Bussey HJR, Morson BC. 1975. The evolution of cancer of the colon and rectum. Cancer 36:2251–2270.
82. Hill MJ, Morson BC, Bussey HJB. 1978. Aetiology of adenoma–carcinoma sequence on the large bowel. Lancet i:245–247.
83. Winawer SJ, O'Brien MJ, Waye JD, Kronborg O, Bond J, Fruhmorgen P, Sobin LH, Burt R, Zauber A, Morson B, and the WHO Collaborating Centre for the Prevention of Colorectal Cancer. 1990. Risk and surveillance of individuals with colorectal polyps. Bull WHO 68:789–795.
84. Bronner MP, Haggitt RC. 1993. The polyp–cancer sequence. Do all cancers arise from benign adenomas? In Sivak MV (ed.), Gastrointestinal Endoscopy. Clinics of North America, vol. 3. Philadelphia: W.B. Saunders Co., pp. 611–622.
85. Dukes CE. 1925. Simple tumors of the large intestine and their relationship to cancer. Br J Surg 13:720.
86. Jass JR. 1989. Do all colorectal carcinomas arise in preexisting adenomas? World J Surg 13:45–51.
87. Lev R. 1990. Adenomatous Polyps of the Colon. Pathobiological and Clinical Features. New York: Springer-Verlag.
88. Tierney RP, Ballantyne GH, Modlin IM. 1990. The adenoma to carcinoma sequence. Surg Gynecol Obstet 171:81–94.
89. Williams CB, Bedenne L. 1990. Management of colorectal polyps: is all the effort worthwhile? J Gastroenterol Hepatol 1 (Suppl):144–165.
90. Eide TJ. 1990. Natural history of adenomas. World J Surg 15:3–6.
91. Itzkowitz SH. 1992. The adenomatous polyp. Semin Gastrointestinal Dis 3:3–12.
92. Neugut AI, Jacobson JS, DeVivo I. 1993. Epidemiology of colorectal adenomatous polyps. Cancer Epidemiol Biomarkers Prev 2:159–176.
93. Grinnell RS, Lane N. 1958. Benign and malignant adenomatous polyps and papillary adenomas of the colon and rectum: an analysis of 1856 tumors in 1335 patients. Int Abstr Surg 106:519–538.
94. Morson BC. 1974. Evolution of cancer of the colon and rectum. Cancer 34:845–849.
95. Fung CHK, Goldman H. 1970. The incidence and significance of villous changes in adenomatous polyps. Am J Clin Pathol 53:21–25.
96. Armitage NCM. 1991. Intervention studies in adenoma patients. World J Surg 15:29–34.
97. Hoff G, Vatn MH. 1991. Colonic adenoma: natural history. Dig Dis 9:61–69.
98. Atkin WS, Morson BC, Cuzick J. 1992. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. N Engl J Med 326:658–662.
99. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AMM, Bos JL. 1988. Genetic alterations during colorectal-tumor development. N Engl J Med 319:525–532.
100. Kuramoto S, Oohara T. 1988. Minute cancers arising de novo in the human large intestine. Cancer 61:829–834.
101. Shimoda T, Ikegami M, Fujisaki J, Matsui T, Aizawa S, Ishikawa E. 1989. Early colorectal carcinoma with special reference to its development de novo. Cancer 64:1138–1146.
102. Bedenne L, Faivre J, Boutron MC, Piard F, Cauvin JM, Hillon P. 1992. Adenoma–carcinoma

sequence or "de novo" carcinogenesis? A study of adenomatous remnants in a population-based series of large bowel cancers. Cancer 69:883–888.
103. Kuramoto S, Oohara T. 1995. How do colorectal cancers develop? Cancer 75:1534–1538.
104. Wolff WI, Shinya H. 1973. Polypectomy via the fiberoptic colonoscope: removal of neoplasms beyond the reach of the sigmoidoscope. N Engl J Med 288:329–332.
105. Kozuka S, Nogaki M, Ozeki T, Masumori S. 1975. Premalignancy of the mucosal polyp in the large intestine. II. Estimation of the periods required for malignant transformation of mucosal polyps. Dis Colon Rectum 18:494–500.
106. Helwig EB. 1947. The evolution of adenomas of the large intestine and their relation to carcinoma. Surg Gynecol Obstet 84:36–49.
107. Correa P, Strong JP, Reif A, Johnson WD. 1977. The epidemiology of colorectal polyps. Prevalence in New Orleans and international comparisons. Cancer 39:2258–2264.
108. Blatt LJ. 1961. Polyps of the colon and rectum: incidence and distribution. Dis Colon Rectum 4:277–282.
109. Offerhaus GJA, Giardiello FM, Tersmette KWF, Mulder GWR, Tersmette AC, Moore GW, Hamilton SR. 1991. Ethnic differences in the anatomic location of colorectal adenomatous polyps. Int J Cancer 49:641–644.
110. Arminski TC, McLean DW. 1964. Incidence and distribution of adenomatous polyps of the colon and rectum based on 1000 autopsy examinations. Dis Colon Rectum 7:249–261.
111. Stemmerman GN, Yatani R. 1973. Diverticulosis and polyps of the large intestine: a necropsy study of Hawaiian Japanese. Cancer 31:1260–1270.
112. Eide TJ, Stalsberg H. 1978. Polyps of the large intestine in Norther Norway. Cancer 42:2839–2848.
113. Vatn MH, Stalsberg H. 1982. The prevalence of polyps of the large intestine in Oslo: an autopsy study. Cancer 49:819–825.
114. Williams AR, Balasovriya BAW, Day DW. 1982. Polyps and cancer of the large bowel: a necropsy study in Liverpood. Gut 23:835–842.
115. Coode PE, Chan KW, Chan YT. 1985. Polyps and diverticula of the large intestine: a necropsy survey in Hong Kong. Gut 26:1045–1048.
116. Bremmer CG, Ackerman LV. 1970. Polyps and carcinoma of the large bowel in the South African Bantu. Cancer 26:991–999.
117. Correa P, Duque E, Cuello C, Haenszel W. 1972. Polyps of the colon and rectum in Cali, Columbia. Int J Cancer 9:86–96.
118. Cajucom CC, Barrios GC, Gruz L, Varin C, Herrera L. 1992. Prevalence of colorectal polyps in Filipinos. An autopsy study. Dis Colon Rectum 35:676–680.
119. Eide TJ. 1986. The age-, sex-, and site-specific occurrence of adenomas and carcinomas of the large intestine within a defined population. Scand J Gastroenterol 21:1083–1088.
120. Johannsen LGK, Momsen O, Jacobsen NO. 1989. Polyps of the large intestine in Aarhus, Denmark. An autopsy study. Scand J Gastroenterol 24:799–806.
121. Rickert RR, Auerbach O, Garfinkel L, Hammond EC, Frasca JM. 1979. Adenomatous lesions of the large bowel. An autopsy survey. Cancer 43:1847–1857.
122. Eide TJ. 1983. Remnants of adenomas in colorectal carcinomas. Cancer 51:1866–1872.
123. DeJong WW, Day NE, Muir CS, Barclay THC, Bras G, Foster FH, Jussawalla DJ, Kurihara M, Linden G, Martinez I, Payne PM, Pederson E, Ringertz N, Shanmugaratnum T. 1972. The distribution of cancer within the large bowel. Int J Cancer 10:463–477.
124. Clark JC, Collan Y, Eide TJ, Esteve J, Ewen S, Gibbs NM, Jensen OM, Koskela E, McLennan R, Simpson JG, Stalsberg H, Zaridze DG. 1985. Prevalence of polyps in an autopsy series from areas with varying incidence of large bowel cancer. Int J Cancer 36:179–186.
125. Rider JA, Kirsner JB, Moeller HC, Palmer WL. 1954. Polyps of the colon and rectum. Their incidence and relationship to carcinoma. Am J Med 16:555–564.
126. Shinya H, Wolff WI. 1979. Morphology, anatomic distribution and cancer potential of colonic polyps. An analysis of 7000 polyps endoscopically removed. Ann Surg 190:679–683.

127. Meltzer SJ, Ahnen DJ, Battifora H, Yokota J, Cline MJ. 1987. Protooncogene abnormalities in colon cancers and adenomatous polyps. Gastroenterology 92:1174–1180.
128. Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759–767.
129. Baker SJ, Preisinger AC, Jessup JM, Paraskeva C, Markowitz S, Willson JKV, Hamilton S, Vogelstein B. 1990. p53 gene mutations occur in combination with 17p allelic deletions as late events in colorectal tumorigenesis. Cancer Res 50:7717–7722.
130. Bonnelli U, Martines H, Conio M, Bruzzi P, Aste H. 1986. Family history of colorectal cancer as a risk factor for benign and malignant tumors of the large bowel: a case-control study. Int J Cancer 41:513–517.
131. Guillem JG, Forde KA, Treat MR, Neugut AI, O'Toole KM, Diamond BE. 1992. Colonoscopic screening for neoplasms in asymptomatic first-degree relatives of colon cancer patients. Dis Colon Rectum 35:523–529.
132. Burt RW, Bishop DT, Cannon LA, Dowdle MA, Lee RG, Skolnick MH. 1985. Dominant inheritance of adenomatous colonic polyps and colorectal cancer. N Engl J Med 312:1540–1544.
133. Cannon-Albright LA, Skolnick MH, Bishop DT, Lee RG, Burt RW. 1988. Common inheritance of susceptibility to colonic adenomatous polyps and associated colorectal cancers. N Engl J Med 319:533–537.
134. McConnell JC, Nizin JS, Slade MS. 1990. Colonoscopy in patients with a primary family history of colon cancer. Dis Colon Rectum 33:105–107.
135. Sauer J, Hoff G, Hausken T, Bjorkheim A, Foerster A, Mowinckel P. 1992. Colonoscopic screening examination of relatives of patients with colorectal cancer. Scand J Gastroenterol 27:667–672.
136. Orrom WJ, Brzezinski WS, Wiens EW. 1990. Heredity and colorectal cancer. A prospective community based, endoscopic study. Dis Colon Rectum 33:490–493.
137. Gryska PV, Cohen AM. 1987. Screening asymptomatic patients at high risk for colon cancer with full colonoscopy. Dis Colon Rectum 30:18–20.
138. Grossman S, Milos ML. 1988. Colonoscopic screening of persons with suspected risk factors for colon cancer. Gastroenterol 94:395–400.
139. Carpenter S, Broughton M, Marks CG. 1995. A screening clinic for relatives of patients with colorectal cancer in a district general hospital. Gut 36:90–92.
140. Gaglia P, Atkins WS, Whitelaw S, Talbot IC, Williams CB, Northover JMA, Hodgson SV. 1995. Variables associated with the risk of colorectal adenomas in asymptomatic patients with a family history of colorectal cancer. Gut 36:385–390.
141. Neugut AI, Jacobsen JS, Ahsan H, Santos J, Garbowski GC, Forde KA, Treat MR, Waye J. 1995. Incidence and recurrence rates of colorectal adenomas: a prospective study. Gastroenterol 108:402–408.
142. Cannon-Albright LA, Bishop DT, Samowitz W, DiSario JA, Lee R, Burt RW. 1994. Colonic polyps in an unselected population: prevalence, characteristics and associations. Am J Gastroenterol 89:827–831.
143. Eide TJ. 1986. Risk of colorectal cancer in adenoma-bearing individuals within a defined population. Int J Cancer 38:173–176.
144. Henry LG, Condon RE, Schulte WJ, Aprahamian C, DeCossee JJ. 1975. Risk of recurrent colon polyps. Ann Surg 182:511–515.
145. Winawer SJ. 1991. Follow-up after polypectomy. World J Surg 15:25–28.
146. Morson BC. 1984. The evolution of colorectal carcinoma. Clin Radiol 35:425–431.
147. Neugut AI, Johnsen CM, Forde KA, Treat MR. 1985. Recurrence rates for colorectal polyps. Cancer 55:1586–1589.
148. Krönborg O, Fenger C. 1987. Prognostic evaluation of planned follow-up of patients with colorectal adenomas. Int J Colorect Dis 2:203–207.
149. Jorgensen OD, Krönborg O, Fenger C. 1993. The Funen Adenoma Follow-up Study. Characteristics of patients and initial adenomas in relation to severe dysplasia. Scand J Gastroenterol 28:239–243.

150. Eddy DM. 1990. Screening for colorectal cancer. Ann Intern Med 113:373–384.
151. Connor RJ, Prorok PA, Weed DL. 1991. The case–control design and the assessment of the efficacy of cancer screening. J Clin Epidemiol 44:1215–1221.
152. Weiss NS, McKnight B, Stevens NG. 1992. Approaches to the analysis of case–control studies of the efficacy of cancer screening. Am J Epidemiol 135:817–823.
153. Moss SM. 1991. Case-control studies of screening. Int J Cancer 20:1–6.
154. Atkin WS, Cuzick J, Northover JMA, Whynes DK. 1993. Prevention of cancer by once-only sigmoidoscopy. Lancet 341:736–740.
155. Gilbertsen VA. 1974. Proctosigmoidoscopy and polypectomy in reducing the incidence of rectal cancer. Cancer 34:936–939.
156. Gilbertsen VA, Nelms MJ. 1978. The prevention of invasive cancer of the rectum. Cancer 41:1137–1139.
157. Miller AB. 1988. Review of sigmoidoscopy screening for colorectal cancer. In Chamberlain J, Miller AB (eds.), Screening for Gastrointestinal Cancer. Toronto: Hans Huber, pp. 3–7.
158. Cutler JL, Ramcharon S, Feldman R, Siegelaub AB, Campbell BA, Friedman GD, Dales LG, Collens MF. 1973. Multiphasic checkup evaluation study I. Methods. Prev Med 2:197–206.
159. Dales LG, Friedman GD, Ramcharon S, Siegelaub AB, Campbell BA, Feldman R, Collen MF. 1973. Multiphasic checkup evaluation study. Outpatient clinic utilization, hospitalization and mortality experience after seven years. Prev Med 2:221–235.
160. Friedman GD, Collen MF, Fireman BH. 1986. Multiphasic health checkup evaluation: a 16-year follow-up. J Chron Dis 39:453–463.
161. Selby JV, Friedman GD, Collen MF. 1988. Sigmoidoscopy and mortality from colorectal cancer: The Kaiser Permanente Multiphasic Evaluation Study. J Clin Epidemiol 41:427–434.
162. Selby JV, Friedman GD, Quesenberry CP, Weiss NS. 1992. A case–control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med 326:653–657.
163. Chapman I. 1963. Adenomatous polypi of large intestine. Incidence and distribution. Ann Surg 157:223–226.
164. Sato E, Ouchi A, Sasano N, Ishidate T. 1976. Polyps and diverticulosis of large bowel in autopsy population of Akita prefecture compared with Miyagi. High risk for colorectal cancer in Japan. Cancer 37:1316–1321.
165. Bombi JA. 1988. Polyps of the colon in Barcelona, Spain. An autopsy study. Cancer 66:1472–1476.
166. Cooper GS, Yuan Z, Landefeld S, Johanson JF, Rimm AA. 1995. A national population-based study of incidence of colorectal cancer and age. Implications for screening in older Americans. Cancer 75:775–781.
167. Lieberman D. 1993. Screening colonoscopy. Has the time come? In Swak MV (ed.), Gastrointestinal Endoscopy Clinics of America. Philadelphia: WB Saunders vol. 3:673–682.
168. Rogge JD, Elmore MF, Mahoney SJ, Brown ED, Troiano FP, Wagner DR, Black DJ, Pound DC. 1994. Low-cost, office-based screening colonoscopy. Am J Gastroenterol 89:1775–1780.
169. Blackburn GL, Giardiello FM. 1995. Developing strategies for intervention/prevention trials of individuals at risk of hereditary colon cancer. Natl Cancer Inst Monogr 17:107–110.
170. Tahara E. 1995. Genetic alterations in human gastrointestinal cancers. The application to molecular diagnosis. Cancer 75:1410–1417.
171. Blum HE. 1995. Colorectal cancer: future population screening for early colorectal cancer. Eur J Cancer 31A:1369–1372.

6. Advances in screening for breast cancer

Sue M. Moss

1. Introduction

By 1990, there was a considerable body of evidence, both from randomized trials and observational studies, that breast screening by mammography in women aged 50 and over was effective in reducing mortality from the disease [1]. It appeared that physical examination provided relatively little additional benefit, while the effect of routine breast self-examination remained unclear.

This chapter reviews evidence that has emerged since then, both on the effectiveness of screening in women over 50 and on the controversial area of screening in women under 50. The current evidence on the value of physical examination and breast self-examination (BSE) is also assessed. Population screening is now becoming policy in many countries, and the status of such screening is described. Potential future developments are discussed briefly.

2. Further evidence of benefit

2.1. Results from Sweden

In 1989 the Swedish two-county study showed a reduction of 32% in women aged 40–74 invited for screening by single-view mammography every 24–33 months [2]. Other trials in Sweden had slightly different protocols in terms of screening interval, age range, and number of views, and the Malmo study had failed to show a significant reduction in women aged 45–69 with a screening interval of 18 to 24 months and using two views at the first two screening rounds [3].

In 1991, preliminary results were published from the Stockholm trial [4], in which women aged 40–64 in the study group were invited to screens by single-view mammography on average 28 months apart. An overall reduction in breast cancer mortality of 29% was observed in the study group (RR 0.71;

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

95% CI, 0.2–1.2); for women aged 50–64 the RR was 0.57 (95% CI, 0.3–1.1), while there was no observed benefit in women under 50 (RR 1.09; 95% CI, 0.3–3.0).

In 1993, an overview of the five Swedish randomized trials was performed [5]. This included data from the Gothenburg trial, from which mortality data had not previously been published. Prior to this analysis, the cause of death in all fatal breast cancer cases was reviewed by an independent endpoint committee [6]. Since in most of the trials the control groups had subsequently been invited for screening, one analysis excluded breast cancer deaths in cases diagnosed after this date, in order to minimize dilution of the effect of screening. This analysis showed consistent relative risks of between 0.68 and 0.84 in the study groups of the five trials, with an overall relative risk of 0.74 (95% CI, 0.66–0.87). When the data were subdivided by age at entry, the relative risk in women aged 40–49 was 0.87 (95% CI, 0.63–1.20), with a reduction of 29% in women aged 50–69, but little evidence of benefit in women aged 70–74 (RR 0.94; 95% CI, 0.60–1.46).

2.2. Results from U.K.

Also in 1993, the results of the 10-year follow-up of the U.K. Trial of Early Detection of Breast Cancer (TEDBC) were published [7]. This was a nonrandomized trial comparing different interventions in different geographical districts. One of the screening districts was identical to the study arm of the Edinburgh randomized trial, which had shown a nonsignificant 17% mortality reduction at seven years [8], but which was hampered by bias in randomization [9]. The U.K. TEDBC showed a 20% reduction at 10 years in women aged 45–64 at entry offered annual screening by mammography and/or physical examination (RR 0.80; 95% CI, 0.67–0.95); the observed benefit in women aged 45–49 did not differ significantly from that in other age groups. In two districts in which women were offered education about BSE, there was no mortality reduction in the combined districts, but a significant difference between the two, which is not completely explained by known differences in treatment policy [10]. The Edinburgh randomized study also reported an 18% mortality reduction in the study group after 10 years of follow-up [11] (RR 0.82; 95% CI, 0.61–1.11).

2.3. The Canadian Breast Screening Study

The Canadian National Breast Screening Study included different protocols for two age groups [12]. In the 50–59 age group, 39,405 women were randomly assigned either to annual mammography and physical examination, or to annual physical examination only. In the 40–49 age group, 50,430 women were randomly assigned to undergo either annual mammography and physical examination, or to usual care after an initial physical examination. All women in both age groups were taught breast self-examination. The study was carried

out in 15 centers in Canada. Women were recruited by various means, the majority volunteering as a result of general publicity.

In 1992, mortality results were published after an average of 8.3 years follow-up in the older women and 8.5 years in the younger age-group [13]. In the 50–59 age group, acceptance rates for screens after the first one were over 85%. There was no difference in breast cancer mortality between the two groups (RR 0.97; 95% CI, 0.62–1.52), despite a cancer detection rate of 7.20 per 1000 in the mammography group at first screen compared with a rate of 3.45 per 1000 in the group offered physical examination only. In the 40–49 age group, over 90% of women attended for screening in years 2 to 5. There was a nonsignificant increase of 36% in breast cancer mortality in the group offered annual screening (RR 1.36; 95% CI, 0.84–2.21).

2.4. Controversy surrounding the Canadian trial

The results of the Canadian trial have aroused considerable debate, particularly those from the study in women aged 40–49, which was the first trial designed specifically to study the effect of screening in women under 50. A number of criticisms of the trial have been expressed concerning both the study design and the quality of screening [14].

One problem, acknowledged by the trial investigators [15], is the lack of statistical power. Both arms of the trial were originally planned to have 80% power to show a reduction of 40% in breast cancer mortality after five years of follow-up. In the event, despite follow-up being extended because of lower than expected mortality in the control group, the mortality rate in the 40–49 control group was only 111 per 100,000 after seven years, compared with 212 per 100,000 after five years, which had been estimated from population mortality rates.

Reservations have also been expressed concerning the quality of the mammography used at some of the centers [16]. Although it was claimed that "almost 50% of mammograms during the first two years of screening were judged unsatisfactory," only a fairly small percentage of all mammograms in the study were performed in this period. In all, some 18% to 24% of mammograms in a sample selected for review were rated unsatisfactory [17], although the sample size and validity of this review have been questioned. Nevertheless, there did appear to be an improvement with mammography within the trial over time.

An excess of advanced breast cancers (with four or more positive nodes) was observed in women in the mammography group compared with the controls, and this has led to some doubts being raised about the randomization process [14], which was carried out after the initial physical examination. However, other data show no evidence of randomization bias, and there appears to have been no incentive for the clinician involved to ensure that a woman with symptoms was allocated to the mammography group, since such women would in any case be sent for review [15]. The suggestion has been

made that the analysis should be repeated excluding all cases diagnosed at the initial physical examination [18].

2.5. *Recent meta-analyses of screening*

Because of differences in study design between the Canadian and other trials, a number of meta-analyses of breast screening trials published in the last few years have produced separate analyses, both including and excluding the Canadian results.

Such analyses have therefore largely added the HIP and Edinburgh trials to the Swedish data, and it should be noted that both former trials included physical examination as a screening modality in the study group.

Elwood et al. [19] conducted a meta-analysis, primarily of the results in women under 50 from five randomized trials at seven years of follow-up (excluding both the Canadian and Gothenburg trials), and found that the death rates in intervention and control groups were virtually identical (RR 0.99; 95% CI, 0.74–1.32) compared with a 34% reduction in the intervention arms in women aged 50–74 (RR 0.66; 95% CI, 0.55–0.79).

Wald et al. [20] carried out an analysis that included updated data from the Edinburgh trial and found an overall relative risk from seven trials of 0.78 (95% CI, 0.70–0.87) for women aged 40–74 at entry. For women aged 50–74, the relative risk was 0.76 (95% CI, 0.69–0.88), and for women aged 40–49 a reduction of 16% was observed (RR 0.84; 95% CI, 0.67–1.06). If the Canadian trial were included, the latter estimate became 7% (RR 0.93; 95% CI, 0.76–1.15).

A more recent meta-analysis, again specifically of women aged 40–49 at entry, included further follow-up from the Gothenburg study [21]. This found a statistically significant reduction of 21% (RR 0.79; 95% CI, 0.64–0.98) if the Canadian study was excluded, and a reduction of 14% (RR 0.86; 95% CI, 0.71–1.05) if it was included.

Another recent meta-analysis [22] included results from case–control studies, including one from the Guildford arm of the U.K. TEDBC, which had been clearly shown to be subject to selection bias [23] when compared to the population-based study, and one from Nijmegen that included women aged 35–39. Including the Canadian study, this analysis found an overall relative risk of 0.75 (95% CI, 0.68–0.83), with a relative risk in women aged 40–49 of 0.93 (95% CI, 0.76–1.13).

Some caution is needed when combining results of randomized trials and case–control studies in this way, since the former are measuring the effect of inviting a population for screening, while the latter compare breast cancer mortality in attenders and nonattenders. However, it is of interest that for women aged 40–49, when only the randomized trials are included, those with follow-up of 10–12 years show a lower relative risk than those with shorter

follow-up, although the difference is not significant. This effect was not observe in older women.

3. Screening in young women

The publication of the results of the Canadian trial has failed to resolve the issue of the effectiveness of breast screening in young women. While the results of the overviews can be viewed as encouraging, it should be borne in mind that these results apply to women up to age 49 at entry, so some of the observed benefit will inevitably be as a result of screening over the age of 50, and two of the trials (Malmo and Edinburgh) only recruited women aged 45 and over. Using the MISCAN simulation model, it has been estimated that 70% of the observed mortality reduction from the Swedish trials in women aged 40–49 at entry was due to screening over age 50 [24]. However, as always, the use of such models requires a number of assumptions, and the predicted mortality reduction for young women, assuming that screening below age 50 has the same effect as in older women, is still within the 95% confidence interval of that observed.

In the United States, the uncertainty has led the National Cancer Institute to change its guidelines for breast screening, which previously had recommended mammography from age 40. In December 1993, the NCI issued a statement that "to date randomized trials have not shown a statistically significant reduction in mortality for women under the age of 50," and removing its support for screening in this age group [25]. The American Cancer Society, meanwhile, has reaffirmed its recommendations for screening, including women aged 40 to 49 [26].

In the United Kingdom, where there is currently less demand for screening from young women and less conviction amongst the medical profession about its efficacy, a large multicenter randomized trial is in progress to attempt to answer this question. The trial aims to compare the effectiveness of starting screening by mammography at ages 40–41 with starting at age 50 as in the U.K. national program. The aim is to recruit 195,000 women, randomized in the proportion 1:2, to a study group offered annual screening and a control group with no intervention.

This trial is designed to have 80% power to show a 20% reduction in breast cancer mortality in the study group over a 10-year period from date of entry. By June 1995, over 124,000 women had been randomized in 21 centers in the U.K., and compliance was running close to the anticipated 70%. An interim analysis is due to be carried out in 1996, in which questions such as the sensitivity of screening, and prognostic factors of breast cancers identified in the trial, will be addressed.

A feasibility study is currently under way to investigate the possibility of setting up a trial with a very similar protocol in four European countries: Italy,

the Netherlands, Sweden, and Finland. Eurotrial 40, if it goes ahead, would commence in 1997.

4. Screening in older women

Results from the Swedish studies suggests that screening is as effective in women aged 65–69 as in those aged 50–64, although there is less evidence of a benefit in women aged 70 and over. In the Swedish overview [6], the relative risk of breast cancer mortality in the study group aged 70–74 was 0.94 (95% CI, 0.60–1.46).

More recently, Chen et al. have shown a reduction of 32% (RR 0.68; 95% CI, 0.51–0.89) in women aged 65–74 [27] in the Swedish two-county study. The relative risk in women aged 65–69 was 0.58 (95% CI, 0.39–0.86), and those aged 70–74 was 0.78 (95% CI, 0.53–1.20). However, the difference between relative risks in the two age groups was not statistically significant, and although attendance was lower in older women, they also received fewer rounds of screening.

The availability of screening to older women is an area of concern in a number of countries. In the U.K., women aged 65 and over are not currently invited for screening, although they may refer themselves. One of the reasons for this decision was a belief that the uptake of screening would be lower among older women. However, a government health committee report has recently concluded that the program should be extended to women aged 65–69 [28]. In the United States, although most guidelines for mammography include women up to at least age 75, surveys have identified a number of barriers to screening in women over age 65, including a lower level of knowledge of the benefit of mammography, lower perceived vulnerability, fewer physician recommendations, and problems of access [29]. These result in lower utilization of mammography in this age group.

5. The effectiveness of physical examination

In the original HIP study, in which screening was carried out both by mammography and physical examination, 45% of cancers detected were reported as being found by clinical examination alone [28].

With subsequent improvement in mammographic quality and technique, the additional benefit of physical examination is generally thought to be small. Results from the screening centers in the U.K. TEDBC, where mammography and physical examination were performed every two years, with physical examination only in the intervening years, suggest the relative sensitivities of mammography and physical examination to be 94% and 70%, respectively, at the first screen, and 91% and 44% at subsequent screens [29]. After adjustment for size and nodal status, the survival of cancers detected at the clinical-

only round was similar to that of the cases in the comparison centers, while that of mammographically detected cancers remained lower [30].

The Canadian study in women aged 50 and over was designed to evaluate the effect of annual mammography over and above that of annual physical examination and the teaching of BSE. Despite the lack of statistical power, the absence of an observed additional benefit suggests a possible beneficial effect of physical examination, but the findings of longer follow-up are required.

Mittra [33] argues the case for screening by physical examination alone, citing the results of the Canadian study and the earlier BCDCP, and also reports that breast cancers found by mammography may have less malignant potential. There is still some debate about the extent to which malignancy of tumors develops over time [34–36]; however, estimates of proportion of cancers with diagnosis advanced by screening still suggest mammography to be the more sensitive test. Nevertheless, the potential of physical examination, particularly in developing countries with limited resources, requires further consideration and evaluation.

6. The effectiveness of breast self-examination

There remains little evidence to support the effectiveness of routine breast self-examination (BSE). In the U.K. TEDBC, two centers invited women aged 45–64 to attend BSE classes and provided open-access clinics for those detecting an abnormality. In an analysis of mortality after 10-year follow-up in one of the centers, Huddersfield showed a reduction in breast cancer mortality similar to that seen in the screening centers, while the other in Nottingham showed a nonsignificant excess [7]. Differences between the centers have been explored [10]; in Huddersfield, 31% of women attended BSE classes as opposed to 53% in Nottingham, but more women in Huddersfield used the special breast clinics set up as part of the program. Attempts to adjust for the greater use of adjuvant therapy in Huddersfield reduced the difference between the two centers only slightly. However, because of the nonrandomized design of the trial, it remains difficult to ascertain what, if any, of the benefit in Huddersfield was due to the BSE program and what was due to unaccounted biases.

In Finland, the Mama program, developed in 1973, identified a population of 56,177 women from selected clubs, who were then offered education and instruction in BSE. They were provided with calendars on which to record BSE practice and findings, and mammography was available as a diagnostic tool. Only those women who returned an initial calendar after two years and who were essentially BSE compliers were included in the analysis. These women's records were linked to the Finnish Cancer Registry.

The analysis, published in 1994 [37], calculated the expected breast cancer mortality in these women using the expected incidence, based on Finnish

Cancer Registry data for 1980, and the expected case–fatality. Breast cancers occurring in the first two years after enrollment were excluded.

A total of 28,785 women were included in the cohort. There was a higher than expected incidence of breast cancer at all ages, although this only appeared from year 9 after entry. The ratio of observed to expected breast cancer deaths was 0.71 (95% CI, 0.57–0.87), with a ratio of 0.64 for women under 50 and 0.74 for those over 50. Although the findings are consistent with the effectiveness of a BSE program in reducing mortality, there is uncertainty due to the possible effect of selection bias, and there is no observed alteration in stage distribution in the study cohort compared with the general population. Also, the fact that much of the reduction in breast cancer mortality occurred in years 3–6 (effectively years 1–4 of cases included) is surprising, given the findings from screening trials that a reduction does not usually begin to appear until years 4 or 5.

Other evidence on BSE comes from case–control studies, which are likely to be affected by selection bias [38]. A study by Newcomb et al. included 209 women members of a Group Health Co-operative who developed advanced breast cancer and 433 age-matched controls from the same population [39]. Cases and controls were interviewed about their BSE practice. The relative risk of advanced breast cancer was 1.15 among "never BSE users" (95% CI, 0.73–1.81), although a small group of women reporting more thorough examinations had a nonsignificant decreased risk. Another study by Muscat and Huncharek [40] shows no difference in BSE frequency reported in cases of advanced breast cancer and population controls, with an odds ratio of 1.27 (95% CI, 0.77–2.07) after adjustment for possible confounding factors. However, the fact that women at higher risk of disease might be more likely to practice BSE makes it difficult to draw firm conclusions from such studies.

A recent case–control study nested within the Canadian NBSS includes as cases 153 breast cancer deaths and 67 cases with distant metastases [41]. Ten controls per case were selected from the trial population, matched by age, center, year, and randomization group. BSE frequency was assessed from annual self-reports, and proficiency from annual assessments by screen examiners were carried out. There was no difference in frequency of BSE practice between cases and controls (RR 1.07 for nonpractitioners; 95% CI, 0.65–1.79), but a significant increased relative risk in those with a low proficiency score compared to a high score (RR 1.76; 95% CI, 1.17–2.65). Only women attending for rescreening are included in this analysis, since others could not have their proficiency assessed.

Ongoing trials in Russia and elsewhere, in which factories and other institutions are randomized, may provide further information. In Moscow and St. Petersburg, a population of over 193,000 women aged 40–64 has been randomized to an intervention group that has been offered BSE education either in groups or on an individual basis. The study is planned to last for 15 years and

to be able to detect a 30% reduction in cumulative breast cancer mortality [42].

Meanwhile, in the U.K., current opinion has moved away from promoting routine breast self-examination towards the concept of "breast awareness," encouraging women to be aware of changes in the breasts and to report any changes early.

7. Population breast screening programs

On the basis of the findings of research trials, a number of countries have now established population-based screening programs. The majority of European programs offer screening by mammography to women aged 50 and over at a two-year interval, although some counties in Finland and Sweden include women from age 40, while in the U.K. a three-year screening interval is being used for women aged 50–64. In addition to the programs in these three countries, a number of pilot projects, funded by the EC, are being set up in various countries in Europe [43].

In the United States, the extent and performance of screening is more difficult to measure. A number of organized programs exist, and women also have access to mammography on a private basis. In Canada, screening is being implemented on a provincial basis.

In New Zealand, two pilot projects are currently in operation in Otago and Waikato, and in Australia a national program is planned, with a two-yearly screening interval and covering women between the ages of 50 and 69; again, screening is being introduced on a pilot basis.

The important outcome measures of the screening performance to be studied include uptake of screening and population coverage, biopsy rate, cancer detection rate, and the rate of detection of small invasive cancers [35]. Because of the differences in the implementation of screening in different countries, some caution needs to be exercised in drawing comparisons between these measures in different countries. The age of women screened will influence the cancer detection rates, for example.

Most national programs began within the last 5 or 6 years and have taken a number of years to cover the total population. It is therefore too early to expect to observe an effect of screening on national breast cancer mortality rates, since in research trials it has taken 4 to 5 years from the start of the trial for a difference to begin to emerge between study and control groups. It is also necessary to remember that in the randomized trials, the reported reduction in breast cancer deaths was in women diagnosed with breast cancer after entry to the trial. For a number of years, population mortality rates will include deaths in cases diagnosed before the introduction of screening, so the effect of screening will be diluted.

In the U.K., the "Health of the Nation" document has set a target of a 25%

reduction in breast cancer mortality in the relevant age group by the year 2000 [45]. However, it has recently been noted that breast cancer mortality rates have already begun to fall in some age groups, probably due to improvements in treatment [46], and this will make estimating the effect of screening more difficult. Declining trends have also been observed in white women in the U.S., where part of the impact may already be due to screening [47].

The implementation of screening will result in an initial rise in incidence rates due to screen-detected cases, and this has already been observed in some countries [48]. Prior to a fall in mortality, one would have to observe a reduction in rates of advanced disease, but unfortunately data are often not available in sufficient detail at a national level.

Another interim evaluation measure is the interval cancer rate, which will give an indication of both the sensitivity of the screening test and the appropriateness of the screening interval. Figures recently published from one region in the U.K. have shown higher than expected rates, particularly in the third year after screening [49], which has led to calls for the screening interval to be reduced. The expected rates, however, have been calculated on the basis of those observed in the Swedish two-county study, and data from other programs also show higher rates than these [50]. Again, there is a need for standardization in the tabulation and definitions used to enable meaningful comparison between countries to be drawn.

8. Advances in mammographic technique

The Swedish two-county study employed single-view mammography, using a single oblique view, and a number of national programs have recommended single-view mammography as the most cost-effective technique. However, many radiologists recommend two-view mammography, and recent evidence from a randomized trial in the U.K. shows that the use of two views at a woman's first screen increases both sensitivity and specificity, increasing the cancer detection rate by approximately 20% and reducing referral rates for assessment by 15% [51]. This result is supported by data from the U.K. national program [52], and the U.K. program will in future include two views at all prevalent screens. The effect of two views at rescreening is not adequately known. The evidence on the effect of double reading of mammograms is also unclear, largely due to the range of protocols used.

Despite recent advances in mammographic techniques, including dose reduction, its accepted limitations have led to investigations of new potential screening methods. Digital mammography is one of the most promising of these. Most studies to date have used conventional mammograms that are then digitized, but further developments should enable the images to be transmitted electronically. Such developments also enhance the potential of computer-aided diagnosis.

8.1. Magnetic resonance imaging

The most promising new technique for breast screening is magnetic resonance imaging (MRI), which provides an in vivo image of soft tissue, with the image contrast resulting from differences in tissue-water proton relaxation times. MRI uses non-ionizing radiation to produce excitation of the protons. The development of tissue-specific contrast agents will improve the sensitivity of the technique. However, its potential use as a screening test is currently also limited by the cost of the equipment involved. Other imaging methods under research include positron emission tomography (PET) and single-photon emission planar CT imaging (SPECT).

9. Disadvantages of screening

Concerns over the radiation hazard from mammography have largely been allayed with the considerable reduction in dose achieved in present-day mammography. However, the ratio between risk and potential benefit will increase with decreasing age of commencing screening [53]. It has also recently been recognized that there may be subgroups of women with increased sensitivity to radiation. One such subgroup is women with the ataxia telangiectasia (AT) gene, who are also believed to be at increased risk of developing breast cancer [54].

The detrimental effect of mammography in increasing the rate of biopsies for benign disease has been much reduced by the increased use of fine needle aspiration (FNA) cytology as a diagnostic tool in assessing mammographically positive cases.

9.1. Anxiety induced by screening

Although it is widely believed that screening can result in increased anxiety due to the invitation to screening or referral for further assessment or biopsy, there is little documentary evidence for this. Ellman et al. [55] found increased anxiety scores in women recalled for assessment in the U.K. TEDBC, but for those with false-positive results, this raised level was relatively short-term, reverting after three months to the same level as in the screened normals, although evidence remained of increased psychological morbidity among breast cancer cases. Bull [56] has attempted to assess the psychological impact of mammography screening and has found no increase in levels of anxiety (as measured by psychometric scores) in four groups of women — those invited for screening, at routine screening, attending for further investigation, and after open biopsy — although increasing levels of BSE practice suggested behavioral changes that might be associated with increased awareness or fear of breast cancer.

9.2. Overdiagnosis

The possibility of overdiagnosis is perhaps the most worrying side effect of mammography screening. In the original HIP study, the numbers of breast cancers in the control and study groups were equal by three years after the end of screening, indicating that overdiagnosis was not a problem [23]. However, the sensitivity of mammography for detecting early lesions has increased, and these will include 15% to 20% of ductal carcinoma in situ [52]. It is evident that not all such cases would progress to invasive cancer, and the appropriate treatment of screen-detected DCIS is the subject of ongoing research.

9.3. Financial cost

A detailed study of the costs of screening carried out in Edinburgh [57] has been updated to 1992/1993 and estimates a cost of £22.60 per women screened (including the costs of assessment) and £4300 per cancer detected [51], while a study in Australia has estimated the same costs as $ Aus 117.7 and 11,550, respectively [58].

The cost-effectiveness of screening will vary between countries, both due to differences in the cost of establishing and running a screening program in different settings and due to differences in background incidence and mortality, which may influence the achievable benefit. A comparison of cost-effectiveness in different EC countries using the MISCAN simulation model has estimated a more than twofold greater cost per life-year gained in Spain than in the Netherlands or U.K. [59].

10. High-risk women

It is estimated that some 5% of breast cancers may have a genetic component. Increasingly, women at a high risk of breast cancer due to a family history of the disease are being offered screening from a young age, and there is a need to monitor the effects of this screening. The recent cloning of the breast cancer genes BRCA1 and BRCA2 are likely to result in demands for testing for gene carriers. When such testing becomes available, it will need extremely careful monitoring, not least because of the uncertainty surrounding the advice to be given to those found positive (see chapter 12, this volume).

11. Conclusions

The evidence that has emerged from breast screening trials over the past five years is consistent with previous estimates of a beneficial effect of screening by mammography in women over age 50. The effect of mammographic screening

below age 50 is still unproven, as is the value of BSE. Ongoing trials may provide answers to these questions.

The effect of physical examination without mammography cannot be measured from existing trials, but would now be difficult to evaluate in many developed countries, where screening by mammography is being introduced on a population basis. Over the course of the next decade, the impact of this mammographic screening on breast cancer mortality should become apparent.

References

1. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC. 1990. Report on a workshop of the UICC Project on evaluation of screening for cancer. Int J Cancer 46:761–769.
2. Tabar L, Fagerberg G, Duffy SW, Day NE. 1989. The Swedish two county trial of mammographic screening for breast cancer: recent results and calculation of benefit. J Epidemiol Commun Health 43:107–114.
3. Andersson I, Aspegren K, Janzon L, et al. 1988. Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. Br Med J 297:943–994.
4. Frisell J, Eklund G, Hellstrom L, Lidbrink E, Rutqvist L-E, Somell A. 1991. Randomised study of mammography — preliminary report on mortality in the Stockholm trial. Breast Cancer Res Treatment 18:49–56.
5. Nystrom L, Rutqvist LE, Wall S, et al. 1993. Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet 341:973–978.
6. Nystrom L, Larsson L-G, Rutqvist L-R, et al. 1995. Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. Acta Oncol 34:145–152.
7. U.K. Trial of Early Detection of Breast Cancer Group. 1993. Breast cancer mortality after 10 years in the U.K. trial of early detection of breast cancer. Breast 2:13–20.
8. Roberts MM, Alexander FE, Anderson TJ, et al. 1990. Edinburgh trial of screening for breast cancer: mortality at seven years. Lancet 335:241–246.
9. Alexander FE, Roberts MM, Lutz W, Hepburn W. 1989. Randomisation by cluster and the problem of social bias. J Epidemiol Commun Health 43:29–36.
10. Ellman R, Moss SM, Coleman D, Chamberlain J. 1993. Breast self-examination programmes in the trial of early detection of breast cancer: ten year findings. Br J Cancer 68:208–212.
11. Alexander FE, Anderson TJ, Brown HK, et al. 1994. The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. Br J Cancer 70:542–548.
12. Miller AB, Howe GR, Wall C. 1981. The National Study of Breast Cancer Screening. 1981. Protocol for a Canadian randomized controlled trial of screening for breast cancer in women. Clin Invest Med 4:227–258.
13. Miller AB, Baines CJ, To T, Wall C. 1992. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. 2. Breast cancer detection and death rates among women aged 50 to 59 years. Can Med Assoc J 147:1459–1488.
14. Kopans DB, Feig SA. 1993. The Canadian National Breast Screening Study: a critical review. AJR 161:755–760.
15. Baines CJ. 1994. The Canadian National Breast Screening Study: a perspective on criticisms. Ann Intern Med 120:326–334.
16. Kopans DB. 1990. The Canadian Screening Program: A different perspective. AJR 155:748–749.
17. Baines CJ, Miller AB, Kopans DB, et al. 1990. Canadian Breast Screening Study: Assessment of technical quality by external review. AJR 155:734–747.

18. Tarone RE. 1995. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 75:997–1003.
19. Elwood JM, Cox B, Richardson AK. 1993. The effectiveness of breast cancer screening by mammography in younger women. On-line J Current Clin Trials 2, Doc. No. 32.
20. Wald NJ, Chamberlain J, Hackshaw A. 1994. European Society of Mastology Consensus Conference on Breast Cancer Screening, Paris, 4–5 February 1993: Report of the Evaluation Committee. Clin Onco l:261–268.
21. Smart CR, Hendrick RE, Rutledge JH II, Smith RA. 1995. Benefit of mammography screening in women aged 40 to 49 years: current evidence from randomized controlled trials. Cancer 75:1619–1626.
22. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. 1995. Effacacy of screening mammography. A meta-analysis. JAMA 273:149–154.
23. Moss SM, Summerley ME, Thomas BT, Ellman R, Chamberlain J. 1992. A case–control evaluation of the effect of breast cancer screening in the United Kingdom trial of early detection of breast cancer. J Epidemiol Commun Health 46:362–364.
24. De Koning HJ, Boer R, Warmerdam PG, et al. 1995. Quantitiative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst 87:1217–1223.
25. Kaluzny AD, Rimer B, Harris R. 1994. The National Cancer Institute and guideline development: lessons from the breast cancer screening controversy. J Natl Cancer Inst 86:901–903.
26. Mettlin C, Smart CR. 1994. Breast cancer detection guidelines for women aged 40 to 49 years: rationale for the American Cancer Society reafirmation of recommendations. Cancer 44:248–255.
27. Chen H-H, Tabar L, Fagerberg G, Duffy S. 1995. The effect of breast screening after age 65. J Med Screening 2:10–14.
28. House of Commons. Health Committee. 1995. Third Report. Breast Cancer Services. Report, together with the Proceedings of the Committee. London: HMSO.
29. Constanza ME. 1994. The extent of breast cancer screening in older women. Cancer 74:2046–2050.
30. Shapiro S, Venet W, Strax P, Venet L. 1988. Periodic screening for breast cancer. In The Health Insurance Plan Project and its Sequelae, 1963–1986. Baltimore & London: Johns Hopkins University Press.
31. Moss SM, Coleman DA, Ellman R, et al. 1993. Interval cancers and sensitivity in the screening centres of the U.K. Trial of Early Detection of Breast Cancer. Eur J Cancer 29A(2):225–258.
32. Moss SM, Ellman R, Coleman D, Chamberlain J, for the United Kingdom Trial of Early Detection of Breast Cancer Group. 1994. Survival of patients with breast cancer diagnosed in the United Kingdom trial of early detection of breast cancer. J Med Screening 1:193–198.
33. Mittra I. 1994. Breast screening: the case for physical examination without mammography. Lancet 343:342–344.
34. Tabar I, Fagerberg G, Day NE, Duffy SW, Kitchin RM. 1992. Breast cancer treatment and natural history: new insights from results of screening. Lancet 339:412–414.
35. Hakama M, Holli K, Isola J, et al. 1995. Aggressiveness of screen-detected breast cancers. Lancet 345:221–224.
36. Day NE, Tabar L, Duffy SW, Chen HH. 1995. Population breast-cancer screening. Lancet 345:853.
37. Gastrin G, Miller AB, To T, et al. 1994. Incidence and mortality from breast cancer in the Mama Program for breast screening in Finland, 1973–1986. Cancer 73:2168–2174.
38. Moss SM. 1991. Case-control studies of screening. Int J Epidemiol 20:1–6.
39. Newcomb PA, Weiss NS, Storer BE, et al. 1991. Breast self-examination in relation to the occurrence of advanced breast cancer. J Natl Cancer Inst 83:260–265.
40. Muscat JE, Huncharek MS. 1991. Breast self-examination and extent of disease: a population-based study. Cancer Detect Prev 15:155–159.

41. Harvey BJ, Miller AB, Baines CJ, Corey PN. In preparation. A case–control study of breast self-examination (BSE) nested in the Canadian National Breast Screening Study.
42. Semiglazov VF, Sagaidak VN, Moiseyenko VM, Mikhaiklov EA. 1993. The Russian Federation/World Health Organisation Study. Study of the role of breast self-examination in the reduction of mortality from breast cancer. Eur J Cancer 29A: 2039–2046.
43. de Waard F, Kirkpatrick A, Perry NM, et al. 1994. Breast cancer screening in the framework of the Europe Against Cancer Programme. Eur J Cancer 3 (Suppl 1):3–5.
44. Day NE, Williams DRR, Khaw KT. 1989. Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer 59:954–958.
45. Department of Health. 1992. The Health of the Nation: A Strategy for Health in England. London: Department of Health, HSMO.
46. Beral V, Hermon, C, Reeves G, Peto R. 1995. Sudden fall in breast cancer rates in England and Wales. Lancet 345:1642–1643.
47. Smigel K. 1995. Breast cancer death rates decline for white women. J Natl Cancer Inst 87(3):173.
48. Feuer EJ, Wun L-M. 1992. How much of the recent rise in breast cancer incidence can be explained by increases in mammography utilization. Am J Epidemiol 136:1423–1436.
49. Woodman CB, Threlfall AG, Boggis CR, Prior P. 1995. Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's northwestern region. Br Med J 310:224–226.
50. Peeters PHM, Verbeek ALM, Hendriks JHCL, Holland R, Mravunac M, Vooijs GP. 1989. The occurrence of interval cancers in the Nijmegen screening programme. Br J Cancer 59:929–932.
51. Wald NJ, Murphy P, Major P, Parkes C, Townsend J, Frost C. 1995. UKCCR multicentre randomized controlled trial of one and two view mammography in breast cancer screening. Br Med J 311:1189–1193.
52. Moss SM, Michel M, Patnick J, Johns L, Blanks R, Chamberlain J. In press. Results from the NHS breast screening programme 1990–1993. J Med Screening.
53. Law J. 1993. Variations in individual radiation dose in a breast screening programme and consequences for the balance between associated risk and benefit. Br J Radiol 66:691–698.
54. Lavin MF, Bennett I, Ramsay J, Gardiner RA, et al. 1994. Identification of a potentially radiosensitive subgroup among patients with breast cancer. J Natl Cancer Inst 86:1627–1634.
55. Ellman R, Angeli N, Christians A, Moss S, Chamberlain J, Maguire P. 1989. Psychiatric morbidity associated with screening for breast cancer. Br J Cancer 60:781–784.
56. Bull AR, Campbell MJ. 1991. Assessment of the psychological impact of a breast screening programme. Br J Radiol 64:510–515.
57. Fraser MN, Clarke PR. 1992. Cost-effectiveness of breast cancer screening. Breast 1:169–172.
58. Hurley SF, Livingstone PM, Thane N, Quang L. 1994. Mammographic screening: measurement of the cost in a population based programme in Victoria, Australia. J Epidemiol Commun Health 48:391–399.
59. van Ineveld BM, van Oortmarssen GJ, de Koning HJ, Boer R, van der Maas. 1993. How cost-effective is breast cancer screening in different EC countries? Eur J Cancer 29A:1663–1668.

7. Prostate cancer screening: current issues

Philip C. Prorok, Arnold L. Potosky, John K. Gohagan,
and Barnett S. Kramer

1. Introduction

Prostate cancer is the most common cancer among males in the United States (excluding skin cancer), and it accounts for more cancer deaths among males than all other cancers except lung cancer. An estimated 244,000 new prostate cancer cases and 40,400 deaths we anticipated in 1995 in the United States [1]. There are no established prevention strategies, although at least one approach using finasteride (Proscar™) is under investigation in a randomized trial [2]. Consequently, many have turned to early detection through screening as a means to control this disease. The potential for this approach has been heightened by the recent development of the prostate-specific antigen (PSA) blood test and transrectal ultrasound (TRUS) to enhance the digital rectal examination (DRE). Although some may believe that one or some combination of these tests is valuable for early detection of prostate cancer, the operating characteristics of these modalities in a screening setting are not well understood, and they have yet to be rigorously evaluated in terms of effect on prostate cancer mortality. It is the purpose of this chapter to describe some of the problems and unresolved issues currently surrounding the evaluation and implementation of screening for prostate cancer.

2. Screening modalities

Early detection of prostate cancer is most effectively accomplished by the combined use of the traditional DRE and the (PSA) serum assay. Transrectal ultrasound has been found useful, not for early detection, but to guide needle biopsies indicated by abnormal DRE or PSA findings and to measure prostate volume, which can be used in an attempt to standardize PSA levels [3].

Prostatic imaging by transrectal ultrasound has been examined by several investigators in observational settings [4–10]. The experience through 1989 was summarized by Waterhouse and Resnick [11]. Sensitivity ranged from 71% to 92% for prostatic carcinoma and 60% to 85% for subclinical disease. Specificity values ranged from 41% to 79% and positive predictive values in

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

the 30% range have been reported. The sensitivity and positive predictive value for ultrasound appear to be better than those for rectal examination. However, the relatively low specificity, the invasiveness, and the cost of transrectal ultrasound mitigate against its use for routine screening [12].

Cooner et al. [5] reported on 1788 men who underwent a rectal examination, PSA determination (Hybritech assay, normal 4.0 ng/ml or less), and a 7-Mhz ultrasound examination. Biopsies were performed on most of the subjects with positive results on ultrasound, as well as a few other subjects. Lee et al. [6] reported similar data on a series of subjects with positive transrectal ultrasound findings who also had digital rectal examination and PSA determinations (Yang assay, normal 2.6 ng/ml or less). In both studies, only 5% to 7% of prostate cancers would have been missed if ultrasound had not been performed.

Several observational studies have estimated process measures for digital rectal examination, but without appropriate controls; with no adjustment for lead time and length biases, the accuracy of these estimates is unknown [4,13]. Investigators have reported that the proportion of clinically localized disease is high when detection is by routine rectal examination [14–16]. However, 50% of digitally detected clinical stage B cases were upstaged to surgical stage C (local tumor extension beyond the prostate capsule) or D1 (metastasis to pelvic lymph nodes) in one series [14]. Resnick [17] summarized the results on rectal examination for detection of prostate cancer in 1987: sensitivity 55% to 69%, specificity 89% to 97%, positive predictive value 11% to 26%, and negative predictive value 85% to 96%. These results are consistent with a more recent review [3]. Data from the ongoing American Cancer Society National Prostate Cancer Detection Project (ACSNPCDP) are also consistent with these statistics: sensitivity 50%, specificity 94%, and positive predictive value 24% [18]. However, operating characteristics of the DRE are likely to depend on the experience of the examiner.

In a case–control study of screening digital rectal examination conducted at Kaiser Permanente Medical Care Program, 139 men with metastatic (stage D) prostatic cancer (cases) were matched with an equal number of men without known prostate cancer (controls) [19]. During the 10-year period before initial diagnosis, the two groups were comparable in terms of numbers of digital rectal examinations performed for routine screening (2.45 from cases versus 2.52 from controls), and evaluation of intestinal or rectal symptoms (0.44 in both groups). There was essentially no difference in the relative risk (as estimated from the calculated odds ratio) of metastatic prostatic cancer for men with one or more screening rectal examinations compared with men with none (0.9 odds ratio; 95% confidence interval, 0.5–1.7).

Serum PSA, too, has been assessed in several observational studies, for initial diagnosis of disease and as a monitoring assay after initial therapy [4–6,20–24]. Sensitivity in the range of 70% and positive predictive values of 26% to 52% have been reported. Sensitivity, specificity, and positive predictive

values for PSA alone estimated from data reported from the ACSNPCDP are 67%, 97%, and 43%, respectively, and for PSA or digital rectal examination 84%, 92%, and 28%, respectively [18]. Variations in practice circumstances and characteristics of the population screened probably explain the wide variability reported.

Clinical algorithms for interpreting screening PSA results are still evolving. A threshold serum PSA value of 4 ng/mL was established by comparing PSA values for 319 patients with organ-confined prostate cancer with values for 597 patients with benign prostatic hyperplasia (BPH). Oesterling [25] summarized the findings as follows: (1) 75% of BPH patients had PSA levels less than 4 ng/mL, whereas only 43% of cases of organ-confined prostate cancer assayed in this range, (2) only 3% of BPH cases assayed at 10 ng/mL or greater, while 20% of the cancers did, and (3) in the intermediate range (4–10 ng/mL), there was substantial overlap — 22% of BPH and 37% of prostate cancers. Clearly, PSA alone could not be relied upon to accurately discriminate between BPH and organ-confined prostate cancer. A threshold of 4 ng/mL was selected as the clinical standard. A recent reassessment based on larger data sets and mathematical modeling techniques has demonstrated that, although a cut-off point of 3 ng/mL would be somewhat more sensitive, the standard of 4 ng/mL is practically equivalent [26].

PSA levels were found to increase as the prostate enlarges, and investigators have sought to standardize PSA measurements against age and size using the ratio of serum PSA to TRUS-determined prostate volume (PSAD) as a discriminant, so as to compensate. There have been reports that the rate of rise in serial PSA values (the so-called PSA velocity or PSAV), use of PSAD, or age-specific PSA cut-off values improve the predictive value of PSA [27]. Carter et al. [24], who conducted a study based on retrospective PSA assays of serially acquired blood for 18 patients over many years, found that PSA levels rise especially rapidly among men with prostate cancer compared to other men. The authors hypothesized that specificity could be improved by serial screens compared to the traditional approach of a fixed threshold. Unfortunately, PSA levels rise and fall dramatically in both benign and early malignant disease, making the selection of a critical rate of change, or PSAV threshold, problematic. Smith and Catalona [28] evaluated 982 serially screened men who initially screened normal using PSA. All underwent biopsy. In this retrospective comparison, the ideal cut-off point varied with age. Smith and Catalona point out that PSA-level variations of 10% have been documented over intervals of a few weeks, making short-term changes unreliable. Brawer et al. [29] hypothesized, based on data collected over a one-year period, that a 20% increase in PSA may identify men at significant risk.

Findings from the ACSNPCDP [30] and pilot data from the randomized controlled prostate cancer screening trial in Rotterdam, the Netherlands, where TRUS, DRE, and PSA were applied to all men [31], do not confirm the superiority of PSAD over the standard PSA test with an upper limit of 4 ng/ml.

The Rotterdam group found that limiting TRUS to the screening of men with serum total PSA between 2.0ng/ml and 10.0ng/ml would eliminate 70% of TRUS exams while missing only one of 17 cancers (5.8%) in their series.

Meshref et al. [32] retrospectively reviewed 3234 cases referred for TRUS, including 2429 patients aged 40–79 years who had previous PSAs without known prostate cancer and received DRE and TRUS. In a subgroup of 236 benign cases, PSA increased with age in parallel with enlarging prostate. Eighteen percent of variation in serum PSA was accounted for by prostate volume, while age accounted for only an additional 2%. PSA density was found useful in selecting cases with negative TRUS, negative DRE, and PSA between 4 and 10ng/ml using a PSAD cut-off point of 0.15. However, in two small European studies, conflicting results have been reported. Tiranti et al. [33] found it impossible to define a PSAD cut-off value useful in distinguishing benign from malignant disease in a comparison of 30 BPH cases with 20 prostate cancer cases, while Wolff et al. [34], in comparing 57 BPH cases with 28 prostate cancer cases, concluded that PSAD was superior to absolute PSA in differentiation.

PSA exists in a variety of forms in serum. This reality offers opportunity for development of new and potentially more sensitive and specific assays. On the hypothesis that PSA is differently glycosolated in benign disease and prostate cancer, a number of investigators have assessed the potential of concanavalin A-bound PSA to discriminate between the two conditions, with mixed, but mostly disappointing, results. [35]

Preliminary results of another method, namely, determination of the ratio of free to total serum PSA as a means to distinguish prostate cancer from benign prostatic hyperplasia, have been reported. Some of the work is described in terms of ratios of serum PSA to gamma-seminoprotein [36,37], while some is described in terms of the ratios of free uncomplexed PSA with total serum PSA or PSA complexed with alpha 1-antichymotrypsin (PSA-ACT) [38–40]. It is unclear whether the gamma-seminoprotein first described by Wang et al. [36] in the 1970s and the moiety used in these later studies are one and the same. It has been reported that what is currently referred to by this terminology is a form of PSA [37,41].

Demura et al. [41] report statistically significant discrimination of prostate cancers by both stage and grade using the ratio of serum PSA to gamma-seminoprotein. In the same analysis they report equivalent sensitivity and improved specificity for the ratio as compared to total serum PSA. Christensson et al. [40] demonstrate that PSA–ACT is the predominant form of PSA in serum drawn from patients with BPH but not prostate cancer (as determined by histological evaluation of transurethral resected tissue from the prostate) and patients with cytopathologically or histopathologically confirmed prostate cancer. Free uncomplexed PSA was a minor fraction of the total serum PSA. They determined that free PSA was a larger fraction in BPH than in untreated prostate cancer cases, and that the ratio of free to total PSA is a powerful statistical discriminator between BPH and prostate cancer in the

cases they evaluated. It is too early to know if this technique will be found to be consistently effective.

In a volunteer group that included 63 men with histologically confirmed BPH, 30 with prostate cancer with an enlarged prostate, and 20 with prostate cancer and a normal-sized prostate, it was found that the percentage of free PSA in serum was prostate volume dependent. After standardization by volume, it was concluded that measurement of free PSA improves specificity, retrospectively eliminating the need for 31.3% of the negative biopsies performed, while detecting at least 90% of the cancers. [42]

Although at this writing there is great flux in the state of knowledge of alternative PSA measures as indicators of prostate cancer, the originally selected cut-off point of total serum PSA at least 4 ng/ml has not been shown to be inferior to any of the alternatives so far evaluated.

3. Population-based rates

Although population-based rates are sometimes used to assess whether trends are consistent with a screening effect [43,44], it is impossible to establish the benefits of screening based upon the analysis of such rates. The interpretation of trends in population-based rates is complex and fraught with multiple competing explanations that are often difficult to sort out. However, we may still ask: What insights can be gained by an evaluation of prostate cancer incidence and mortality rates over the last two decades?

From the early 1970s through the mid-1980s, prostate cancer incidence rates in the U.S. increased rapidly, with almost the entire increase in the incidence of localized tumors, while prostate cancer mortality rates remained nearly constant (figure 1). Both in the U.S. and in Canada, the increase during this time period was attributed to the increasing detection of latent, asymptomatic disease from the more frequent use of transurethral resection of the prostate (TURP) for treatment of obstructive symptoms of benign prostatic hyperplasia (BPH) [45,46]. Other investigators reported an association between early detection and increasing prostate cancer incidence rates [47,48]. As a result of the increase in detection of latent prostate cancer, there has been a dramatic improvement in reported stage-specific survival rates during the period when there were no major advances in treatment [49]. This can mostly be attributed to the addition of a large number of nonaggressive tumors to the existing mix of cases, with a resulting increase of lead time and length bias.

In more recent years, there has been a decline in the use of TURP in the U.S. and an increase in the use of needle biopsy of the prostate [50]. Thus, incidental detection has given way to intentional detection of localized disease. The main reason for the increasing rate of needle biopsy of the prostate since 1987 in the U.S. was an exponential increase in PSA screening starting in 1989–1990 (figure 2). The overall age-adjusted rate of prostate cancer doubled in the U.S. from 1986 to 1992, going from 90/100,000 men to 187/100,000 men

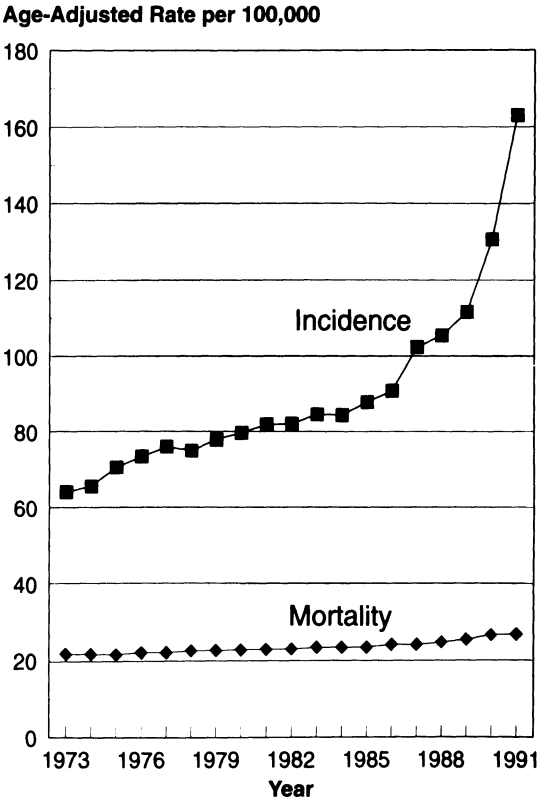


Figure 1. Prostate cancer incidence and mortality trends in the U.S.A. Incidence rates are from the SEER Program, NCI. Mortality rates are from the National Center for Health Statistics.

in 1992. The two single largest annual increases occurred in 1990 and 1991, after PSA had first been introduced and aggressively promoted by several professional organizations. Unlike the case with TURP in the previous decade, the rate of increase in localized incidence rates has been slightly lower than the rise in regional incidence rates. Over the same time, the rate of distant stage prostate cancer has remained essentially constant.

The dramatic increase in incidence rates of both localized and regional-stage prostate cancer since the mid-1980s is the result of two related phenomenon. First, there has been a large increase in the number of new cases due to more frequent screening using PSA, a more sensitive test than physical examination for early prostate cancer. The use of such a test is capable of detecting a substantial number of tumors in men ages 50 and older. The prevalence of latent prostate cancer in this age group is estimated to be approximately 30%, based on autopsy studies [51,52]. The second phenomenon related to trends in localized and regional incidence rates is the increasing use of radical prostatectomy, which began in the early 1980s [53–56]. Since U.S. incidence rates are reported in terms of pathological rather than clinical stage, the

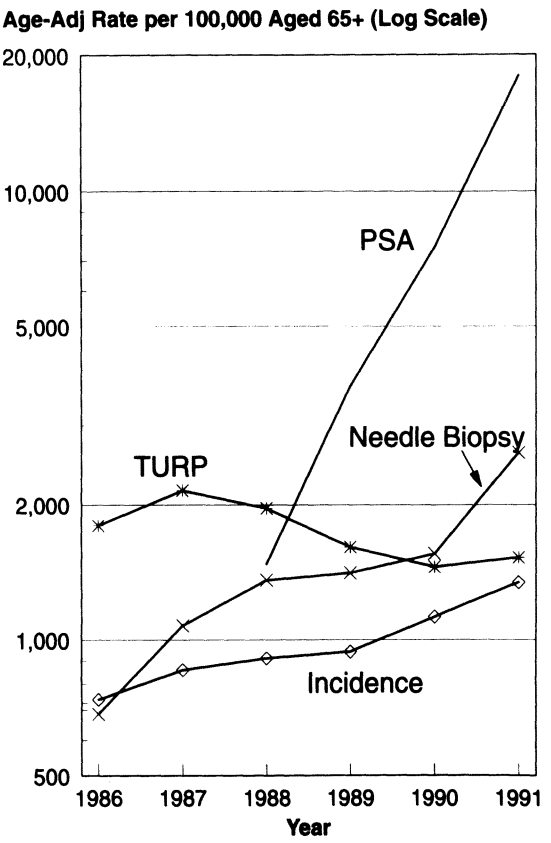


Figure 2. Trends in incidence and diagnostic procedures for prostate cancer among Medicare enrollees in four SEER areas (Detroit, Atlanta, Seattle, Connecticut), 1986–1991. PSA = Prostate Specific Antigen serum test (no Medicare code existed for PSA prior to 1988); TURP = Transurethral Resection of the Prostate; Needle Biopsy = needle biopsy of the prostate.

increasing use of radical prostatectomy has resulted in the more frequent upstaging of clinically localized tumors to pathologically confirmed regional disease. Radical prostatectomy permits the confirmation of the presence of tumor extension beyond the prostate into adjacent structures and pelvic lymph nodes.

Preliminary reports from several individual state tumor registries indicate that the rise in prostate cancer incidence may have begun to reverse in 1993 and 1994 (e.g., [57]). This trend may be due to an overall decline in the use of PSA screening. However, if PSA rates continue to increase, this trend may indicate that PSA is being used for repeat screenings of a subset of men and that there remains a substantial proportion of elderly men who have yet to be screened with PSA.

If the slight decline observed in distant (metastatic-stage) prostate cancer incidence rates since 1991 continues, this may portend a subsequent decline in mortality rates. On the other hand, the decline in distant-stage incidence may

be a reflection of length bias. Cases being detected earlier and thus removed from the pool of distant cases may be disproportionately those cases with the slowest progressing tumors. The remaining distant cases may be the most aggressive tumors. Thus, any decline observed in distant-stage incidence is not necessarily a valid proxy for a reduction in prostate cancer mortality.

Mortality rates from prostate cancer have risen slowly over the last two decades, with the greatest increases observed for men ages 75 and older. If screening were to impart a mortality benefit, several years of observation would be required to detect such an effect, if the effect were large enough to detect in the midst of random variation in population-based rates. In the meantime, the best chance for establishing a mortality benefit remains the completion of ongoing randomized controlled trials of screening.

4. Natural history and overdiagnosis

With the advent of the PSA test, screening for prostate cancer has become more appealing both to the public and to the medical community. The test is simple and inexpensive, and it has higher sensitivity compared to DRE, especially for prostate-confined cancers [20]. As noted above, while the rate of PSA testing has increased dramatically in recent years, so has the incidence of prostate cancer. This increase has been accompanied by recognition of the very real possibility that disease is being diagnosed that would never surface in the absence of screening, as suggested by the high prevalence of latent prostate cancer at autopsy among men aged 50 and over.

It has been claimed that only a small proportion of PSA-detected prostate cancers (stage T1c) are "clinically insignificant"; only 16% of 157 consecutive men in one case series who underwent radical prostatectomy for stage T1c disease had "insignificant" tumors [58]. It was therefore proposed that about 84% of such patients warrant definitive therapy, while watchful waiting might be appropriate for the other 16%. The definition of clinical significance was based upon tumor volume, differentiation, and capsular penetration. However, the definition of clinical significance and the prediction of the clinical behavior of such tumors require substantial caution. These are not based upon actual observed outcomes but rather on comparison to historical series of patients who were diagnosed prior to the implementation of screening. It is difficult to predict the clinical behavior of an asymptomatic cancer picked up only through screening simply by knowing the behavior of a tumor with similar size and histologic characteristics that brought the patient to the doctor because of symptoms [59]. The former is likely to possess a more indolent natural history because of the length bias inherent in the screening process. This has been described in the case of screening DRE [60].

An extreme form of length bias is termed *overdiagnosis*, also referred to as the diagnosis of "pseudodisease" [61]. This can occur when some cases of disease detected by screening are so indolent that they would never have come

to medical attention before the individuals died of other causes if the individuals did not participate in screening. Nevertheless, the cancers are detected through screening and treated, and the disease does not recur. As a result, it appears as though early detection has led to improved outcome, when this may not be the case. A key question is whether overdiagnosis is a realistic possibility with PSA screening.

The best way to document the phenomenon of overdiagnosis is in a randomized trial employing the stop screen design, in which the intervention arm is screened for a limited time period after which both arms are followed for cancer incidence as well as mortality. If an excess of cases persists in the screened arm, this is evidence of overdiagnosis [62]. Randomized screening studies of breast cancer and lung cancer give strong evidence that overdiagnosis exists for both diseases [63,64]. In both studies, there were consistently more cases detected in the screened group than in the control groups. Further, most of the excess cases detected by screening in the lung cancer trial were early-stage surgically operable lesions [64]. Yet the total numbers of deaths from lung cancer were virtually identical in the screened and control arms.

Prostate cancer manifests a wider spectrum of clinical behavior than lung cancer, and most would accept that a large proportion can be indolent. No randomized trials have yet been reported that could document overdiagnosis, but observational studies do raise the possibility. As mentioned above, up to two thirds of men over the age of 80 harbor prostate cancer at autopsy [51,52], but only about 3.4% of American men die of prostate cancer [49]. Thus, far more men die with prostate cancer than die of prostate cancer. The possibility of overdiagnosis is further strengthened by data from a population-based Swedish series of patients with clinically diagnosed stages A or B prostate cancer who were treated by watchful waiting [65]. Only 8.5% of these men had died of their prostate cancer, and only 15% of the deaths in the series were due to prostate cancer after 10 years of follow-up. Even clinical stage C may have an indolent natural history if left untreated. In a case series of 50 patients in Sweden who had extracapsular disease, primarily well or moderately well differentiated, and who were followed on a surveillance protocol, prostate-specific mortality was 12% at five years and 30% at nine years after diagnosis [66].

However, the population "experiment" of prostate cancer screening that is currently going on in the U.S. provides the most compelling evidence of overdiagnosis to date [50]. There has been wide geographic variation in use of PSA for screening in the United States (figure 3). This has resulted in wide geographic variation in the incidence and treatment of prostate cancer. In recent years, for example, there has been more than a 20-fold difference in prostatectomy rates per 100,000 male Medicare beneficiaries between the states of Rhode Island and Alaska [54]. If there were no overdiagnosis, one would expect that mortality would parallel incidence. However, there appears to be no association between incidence and mortality (figure 4) [53]. One could

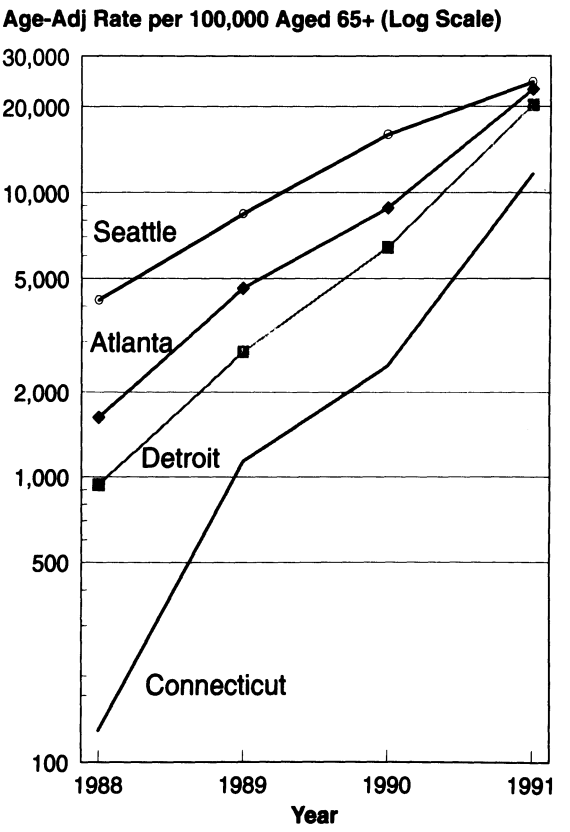


Figure 3. Differences in the use of PSA testing among whites in four SEER areas, 1988–1991. Source: Potosky, Miller, Albertsen, Kramer [50].

postulate that this lack of association is a consequence of an exact match between geographic variations in incidence and in therapeutic efficacy, but this explanation is difficult to accept. Alternatively, in time, mortality may more closely track incidence. However, incidence has been increasing since the early 1970s in the U.S. as early detection modalities have been in increasing use, but mortality rates changed little [45,50].

5. Treatment

Two components must be successful if screening is to work: the screening test must detect the disease early, and treatment initiated at the earlier time point must favorably alter the natural history [67]. It is therefore important to examine what is known about the efficacy of treatment of prostate cancer in any assessment of screening. Favorable survival after definitive treatment of organ-confined prostate cancer is well established. However, survival is the consequence of treatment superimposed upon the natural history of the dis-

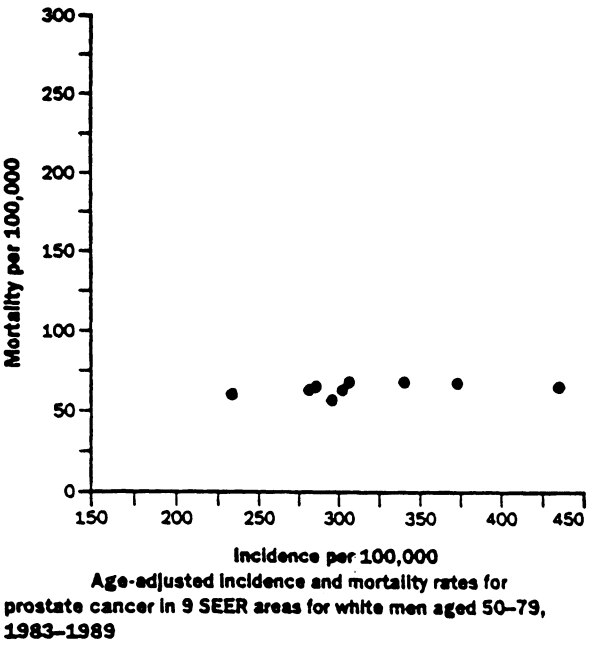


Figure 4. Age-adjusted incidence and mortality rates for prostate cancer in nine SEER areas for white men aged 50–79, 1983–1989. Source: Lu-Yao and Greenberg [53].

ease. Knowledge of the relative contribution of treatment versus natural history to overall survival is crucial to understanding the potential impact of screening. Unfortunately, in prostate cancer, there is insufficient knowledge to accurately assess the relative merits.

The strongest evidence of benefit of any therapy is the demonstration in a randomized controlled trial of prolonged survival after treatment compared to observation of the natural history of the disease. To date, only one small trial has been performed and reported. This trial, begun in 1967, compared radical prostatectomy plus placebo to placebo alone in 111 patients with clinical stages I and II prostate cancer [68,69]. The overall survival rates at 15 years in the two study arms were remarkably similar and approached normal life expectancy. However, the small size and low statistical power of the trial preclude any definitive statements against surgery. At the same time, one cannot conclude that radical prostatectomy is effective, and the study suggests that a large proportion (if not all) of the favorable prognosis of localized prostate cancer is due to the natural history of disease. New trials have recently been undertaken to address the question. The Prostate Intervention Versus Observation Trial (PIVOT) in the U.S. compares radical prostatectomy to palliative expectant management for clinically localized prostate cancer [70]. This study is designed to enroll 2000 patients and is estimated to have a 90% power to detect a 15% reduction in all-cause mortality. A trial in Sweden and Finland is comparing

radical prostatectomy to deferred treatment, and one in Denmark is comparing radiation to watchful waiting [65].

Since the results of these trials will not be available for years, attention has been focused on other forms of evidence. One approach was a structured review of the English language literature spanning 1966 through 1991 that found 1600 articles regarding management of clinically localized prostate cancer [71]. However, all but 144 articles were excluded because they lacked primary data or reported the experience of less than 15 patients. It was noted that little of a definitive nature could be gleaned from the literature review, since the study designs did not permit direct comparison of the three primary treatment strategies, namely, radical prostatectomy, definitive radiation therapy, and expectant management or watchful waiting. The authors concluded that the current state of evidence precludes optimal informed choices by patient or physician.

Several decision models of treatment for localized prostate cancer have been developed to attempt to overcome the state of uncertainty about the value of current treatment options [72–74]. Two of these suggest marginal changes in life expectancy resulting from active therapy, but a net loss in quality-adjusted life expectancy in most situations [72,73]. One suggests considerable gains in both parameters [74]. These models are subject to criticism, however, since they are based on the same literature discussed above, from which it is not possible to draw definitive conclusions. Thus, an important conclusion is that current therapies cannot be evaluated or ranked based upon available data; definitive studies are needed.

The other aspect of therapy about which more is known is the morbidity caused by treatment for localized prostate cancer. The most severe adverse outcome of treatment is mortality, where prostatectomy is associated with a higher risk of treatment-related death than radiation therapy. Some surgeons report mortality rates less than 0.5%, depending upon patient selection and other factors. However, national mortality rates reported for 30-day postoperative mortality are 1% to 2% from a 20% national sample of male Medicare beneficiaries age 65 or older [54]. In addition, about 8% of men suffered major cardiopulmonary complications within 30 days of prostatectomy. In contrast, a recent study of men under 65 years of age reported a 30-day mortality rate of only 0.28% [75].

Other side effects of both radical prostatectomy and radiation therapy include sexual impotence, rectal injury, urinary incontinence, and urethral stricture. The frequency and gravity of these conditions vary depending upon who is reporting the outcomes — the treating physician or the patient. Reports from urologists and radiation therapists give rates of about 25% to 40% impotence, 1% to 3% rectal injury, 3% to 6% urinary incontinence, and 8% to 18% urethral stricture, with radiation therapy in the lower range relative to surgery [18]. However, patient-reported problems are considerably higher. Over 30% of the men reported the need for pads or urinary clamps, and 63% reported a current problem with wetness in a national survey of Medicare

patients who underwent radical prostatectomy in 1988–1990 [76]. About 60% of the men reported having no erections since surgery, and 90% had no erections sufficient for intercourse in the month prior to answering the survey.

Of considerable interest is the morbidity associated with the newer anatomic ("nerve-sparing") prostatectomy procedure. A study at the University of Wisconsin sent the same questionnaire as in the Medicare survey to 93 consecutive patients treated with radical prostatectomy, primarily the anatomic procedure [77]. The reported frequencies of impotence and urinary incontinence were similar to those found in the national survey, despite a younger average age and shorter duration of follow-up in the case series. Additionally, a cross-sectional sruvey of men who had undergone either radical prostatectomy, definitive radiation therapy, or watchful waiting for clinically localized prostate cancer showed substantial sexual and urinary dysfunction in the two active therapy groups, even when controlled for underlying dysfunction experienced by older men without cancer [78]. Moreover, within the context of this small study, sexual function and urinary dysfunction did not differ significantly between men who had undergone standard versus nerve-sparing prostatectomy.

6. Costs and effects of prostate cancer screening

The basic premise that motivates the use of economic analysis to help guide medical decision making is that limited health care resources must be allocated among competing prevention, screening, or treatment interventions (see chapter 3, this volume).

To estimate the net health effect of different types of cancer screening programs, several investigators have employed a cost-effectiveness analysis (CEA), also referred to as clinical decision analysis (CDA).

To date, there have been several CEAs of various strategies of prostate cancer screening using quantitative decision analysis [79–82]. Optenberg and Thompson [79] performed a clinical decision analysis of different screening procedures for prostate cancer among men ages 50–70, including DRE, TRUS, prostatic acid phosphatase, and PSA. They took into account the various potential adverse effects of treatment complications that might occur under a mass screening program. For the most favorable screening programs, the estimated cost per life-year saved was in excess of $200,000. Given the uncertainties about benefits and the known adverse consequences of treatments for early-stage disease, the authors suggested that mass screening for prostate cancer was not an advisable public policy. However, the authors acknowledged that the probabilities employed in their decision analysis pertaining to disease prevalence, disease progression, and test characteristics required further refinements before any definitive conclusions could be drawn about cost-effectiveness of PSA screening.

It is essential that all downstream costs of treatment from diagnosis through

death be incorporated in the cost–benefit calculations. When estimating the costs of screening, Littrup et al. [80] assumed that the cost of treating advanced prostate cancer was $55,000 greater than the cost of treating early-stage cancer, which may have contributed to their observation of a favorable cost–benefit ratio for early-detection strategies. More recent reports demonstrate that the difference in treatment costs according to stage at diagnosis are not substantial over the entire course of the disease and may even be greater for early-stage cases. Two reports provide estimates of stage-specific treatment costs for prostate cancer that might be useful in future estimates of cost-effectiveness of screening. One report, measuring total lifetime payments to Medicare for men ages 65 and over, showed the costs of prostate cancer cases to be $53,000 for localized and regional-stage cases and $33,677 for distant-stage disease [83]. The lower total lifetime costs for distant disease is primarily due to the shorter survival of cases, which, in turn, can be in part an artifact of lead time bias in the screening setting. However, this estimate includes all costs, including those unrelated to the care of prostate cancer. Another group of investigators at a large western Health Maintenance Organization estimated the cancer-related, stage-specific costs for prostate cancer [84]. Rather than providing lifetime estimates, they calculated phase-specific average costs for the initial, continuing, and terminal phases of cancer care. Initial care, comprising the first six months following diagnosis, for localized prostate cancer was $9,300 compared with $8,300 for distant stage disease. Terminal costs during the final six months of life were higher for localized disease at $19,000 versus $11,000 for distant-stage disease. The continuing costs were $1,800 per year for distant stage, compared with $1,200 for localized stage. These results demonstrate that costs vary little by stage at diagnosis.

Another decision analysis focused on screening for prostate cancer using PSA, TRUS, and DRE among men ages 50–70 [81]. These investigators adjusted the estimated gains in life expectancy for potential decrements in quality of life by weighting life-years using the utilities (or preferences) for various health states as the weights. After estimating the costs per quality-adjusted life-year under various screening strategies, the authors concluded that screening prolonged life expectancy but diminshed quality-adjusted life expectancy. All programs increased costs, and results were sensitive to assumptions about the benefits of treating early-stage prostate cancer. Even in high-prevalence populations such as U.S. black males, screening produced losses in quality-adjusted life expectancy and increased costs. However, the authors acknowledged that assessment of comorbidity, risk attitude, and valuation of sexual function may identify some men who might benefit from screening.

Another analysis sponsored by the U.S. Office of Technology Assessment used a quantitative decison model to estimate the risks, benefits, and costs of an early-detection program for a one-time screening program using PSA and DRE for men ages 65, 70, and 75 [82]. The key assumptions to which results were most sensitive, as in other published decision models, concerned the

effectiveness of early treatment and the rate of progression for clinically localized cancers. Under the most favorable assumptions, the costs per life-year saved ranged from $14,200 for men aged 65 to $51,290 for men aged 75. However, assumptions more closely matching available information from the published literature yield estimates of $95,000 at age 65 to $500,000 at age 75. As in the case of the other cost-effectiveness analyses summarized here, the authors point out the need for more research to resolve the uncertainties about the potential benefits of screening so that a definitive cost-effectiveness analysis can be performed.

For the purpose of determining the impact of prostate cancer screening on the health care budget, it is useful to estimate the net cost of a mass screening program. Optenberg and Thompson [79] estimated the total annual cost of prostate cancer screening using a single PSA screening test at $27.9 billion, using a cutoff value of 4 ng/ml. Using more conservative assumptions, the estimated total first-year cost of a hypothetical screening program for men ages 50 to 64 was found to be about $11.9 billion [18].

7. Discussion

Rigorous evaluation of prostate cancer screening modalities is mandatory because the natural history of the disease is not well understood and the effectiveness of treatment of early detected cases is not established. The behavior of individual tumors is unpredictable, and the necessity to treat a particular case of prostate cancer cannot be proved to date. Given the possibility of unnecessary morbidity and mortality associated with diagnosis and treatment of many such lesions, careful evaluation of prostate cancer screening is needed before it is promulgated to asymptomatic men as a public health policy.

The current state of knowledge does not permit a truly informed decision with regard to routine prostate cancer screening and subsequent management. This unfortunate dilemma exists in part because it has taken so long to initiate studies that can test extant assumptions about screening and treatment. Trials are finally under way to fill in the knowledge gaps. Therapy trials of definitive surgery and of radiation versus deferred therapy have been initiated [65,70]. Further, randomized controlled screening trials in the Netherlands, Belgium, Finland, Sweden, Portugal, Italy, and potentially other European countries (the Pan-European Trial), two trials in Canada, and the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial in the United States are evaluating PSA with and without DRE as screening tools. All have prostate cancer mortality as their endpoint. Data from these trials over the next decade will establish the operating characteristics of the modalities for early detection among initially asymptomatic populations, in addition to determining whether repeated screening at different frequencies over a number of years followed by appropriate treatment detects prostate cancer early

enough to effect a significant mortality reduction. An international overview analysis collaboration among screening trial investigators has been initiated [85].

In the meantime, what should be done? In the light of real and potential harm from the screening and therapeutic processes, it does not seem appropriate to encourage screening programs, with their implied promise of benefit. Doing so would deviate from the Hippocratic principle of "first do no harm." Rather, it would seem reasonable for the health care community to inform each man about the current state of uncertainty, describe the risks and theoretical benefits, and encourage participation in ongoing trials when practical. Outside the study setting, screening tests should only be done after a man has been engaged in the decision process as a full, informed partner.

References

1. Wingo PA, Tong T, Bolden S. 1995. Cancer statistics. CA Cancer J Clin 45:8–30.
2. Feigl P, Blumenstein B, Thompson I, et al. 1995. Design of the prostate cancer prevention trial (PCPT). Controlled Clin Trials 16:150–163.
3. Bentvelson FM, Schroder FH. 1995. Modalities available for screening for prostate cancer. Eur J Cancer 29A:811–813.
4. Babaian RJ, Mettlin C, Kane R, Murphy GP, Lee F, Drago JR, Chesley A. 1992. The relationship of prostate-specific antigen to digital rectal examination and transrectal ultrasonography. Findings of the American Cancer Society National Prostate Cancer Detection Project. Cancer 69:1195–1200.
5. Cooner WH, Mosley BR, Rutherford CL, et al. 1988. Clinical application of transrectal ultrasonography and prostate specific antigen in the search for prostate cancer. J Urol 139:758–761.
6. Lee F, Torp-Pedersen S, Littrup PJ, et al. 1989. Hypoechoic lesions of the prostate: clinical relevance of tumor size, digital rectal examination, and prostate-specific antigen. Radiology 170:29–32.
7. Scardino PT. 1989. Early detection of prostate cancer. Adv Urol Ultrasound 16:635–655.
8. Chodak GW, Wald V, Parmer E, Watanabe H, Ohe H, Saitoh M. 1986. Comparison of digital examination and transrectal ultrasonography for the diagnosis of prostatic cancer. J Urol 135:951–954.
9. Lee F, Littrup PJ, Torp-Pedersen ST, et al. 1988. Prostate cancer: comparison of transrectal US and digital rectal examination for screening. Radiology 168:389–394.
10. Miller GJ. 1989. Histopathology of prostate cancer: prediction of malignant behavior and correlation with ultrasonography. Urology 33 (Suppl 6):18–26.
11. Waterhouse RL, Resnick MI. 1989. The use of transrectal prostatic ultrasonography in the evaluation of patients with prostatic carcinoma. J Urol 141:233–239.
12. Pollack HM, Resnick MI. 1993. Prostate-specific antigen and screening for prostate cancer: much ado about something? Radiology 189:353–356.
13. Chodak GW, Schoenberg HW. 1984. Early detection of prostate cancer by routine screening. JAMA 252:3261–3264.
14. Chodak GW, Keller P, Schoenberg HW. 1989. Assessment of screening for prostate cancer using the digital rectal examination. J Urol 141:1136–1138.
15. Donohue RE, Fauver HE, Whitesel JA, Pfister RR. 1979. Staging prostatic cancer: a different distribution. J Urol 122:327–329
16. Thompson IM, Rounder JB, Teague JL, Peek M, Spence CR. 1987. Impact of routine screening for adenocarcinoma of the prostate on stage distribution. J Urol 137:424–426.

17. Resnick MI. 1987. Editorial comments. In Rattiff TL, Catalona WJ (eds.), Genitourinary Cancer. Boston: Martinus Nijhoff Publishers, pp. 94–99.
18. Kramer BS, Brown ML, Prorok PC, Potosky AL, Gohagan JK. 1993. Prostate cancer screening: what we known and what we need to know. Ann Intern Med 119:914–923.
19. Friedman GD, Hiatt RA, Quesenberry CP, Selby JV. 1991. Case–control study of screening for prostatic cancer by digital rectal examinations. Lancet 337:1526–1529.
20. Catalona WJ, Smith DS, Ratliff TL, et al. 1991. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med 324:1156–1161.
21. Lange PH, Ercole CJ, Lightner DJ, Fraley EE, Vessella R. 1989. The value of serum prostate specific antigen determinations before and after radical prostatectomy. J Urol 141:873–879.
22. Oesterling JE, Chan DW, Epstein JI, et al. 1988. Prostate specific antigen in the preoperative and postoperative evaluation of localized prostatic cancer treated with radical prostatectomy. J Urol 139:766–772.
23. Stamey TA, Yang N, Hay AR, et al. 1987. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med 317:909–916.
24. Carter HB, Pearson JD, Metter EJ, et al. 1992. Longitudinal evaluation of prostate-specific antigen levels in men with and without prostate disease. JAMA. 267:2215–2220.
25. Oesterling JE. 1991. Prostate specific antigen: a critical assessment of the most useful tumor marker for adnocarcinoma of the prostate. J Urol 145:907–923.
26. Labrie F, Dupont A, Suburu R, et al. 1992. Serum prostate specific antigen as pre-screening test for prostate cancer. J Urol 147:846–852.
27. Takayama T, Vessella R, Lange P. 1994. Newer applications of serum prostate-specific antigen in the management of prostate cancer. Semin Oncol 21(5):542–553.
28. Smith DS, Catalona WJ. 1994. Rate of change in serum prostate specific antigen levels as a method for prostate cancer detection. J Urol 152:1163–1167.
29. Brawer MK, Beatie J, Wener MH, et al. 1993. Screening for prostate cancer with prostate specific antigen: results of the second year. J Urol 150:106–109.
30. Mettlin C, Littrup PJ, Kane RA, et al. 1994. Relative sensitivity and specificity of serum prostate specific antigen (PSA) level compared with age-referenced PSA, PSA density, and PSA change. Cancer 74:1615–1620.
31. Bangma CH, Grobbee DE, Schroder FH. 1995. Volume adjustment for intermediate prostate-specific antigen values in a screening population. Eur J Cancer 31A:12–14.
32. Meshref AW, Bazinet M, Trudel C, et al. 1995. Roles of prostate-specific antigen density after applying age-specific prostate-specific antigen reference ranges. Urology 45:972–979.
33. Tiranti O, Annesci S, Montefiore F, Boccafoschi C. 1994. Utilia della PSA density (PSAD) nella diagnosi differenziale tra adenocarcinoma prostastatico ed ipertrofia prostatica benigna.: Arch Ital Urol LXVI (Suppl 4):59–63.
34. Wolff JM, Scholz A, Boeckmann W, Jakse G. 1994. Differentiation of benign prostatic hyperplasia and prostate cancer employing prostatic-specific antigen density. Eur Urol 25:295–298.
35. Marrink J, Klip H, De Jong R. 1992. Prostate-specific-antigen-con A binding ratio in benign prostate hyperplasia and prostate cancer (letter to the editor). Lancet 339:619–620.
36. Wang MC, Papsidero LT, Chu TM. 1994. Prostate-specific antigen, p30, gamma-seminoprotein, and E1. Prostate 24:107–108.
37. Demura T, Watarai Y, Togashi M, et al. 1993. Measurement of prostate specific antigen and gamma-seminoprotein ratio: a new means of distinguishing benign prostatic hyperplasia and prostate cancer. J Urol 150:1740–1745.
38. Bangma CH, Kranse R, Blijenberg BG, Schroder FH. 1995. The new Delphia prostate specific anttigen {sic} (PSA) assays of free and total PSA: results of the first comparative evaluation. Proc Am Urol Assoc 153 (April 1995 Suppl):294A.
39. Partin AW, Kelly CA, Subong ENP, et al. 1995. Measurement of the ratio of free PSA to total PSA improves prostate cancer detection for men with total PSA levels between 4.0 and 10.0ng/ml. Proc Am Urol Assoc 153 (April 1995 Suppl):295A.

40. Christensson A, Bjoerk T, Nilsson O, et al. 1993. Serum prostate specific antigen complexed to alpha 1-antichymotrypsin as an indicator of prostate cancer. J Urol 150:100–105.
41. Demura T, Ohyama I, Togashi M, et al. 1995. Prostate specific antigen (PSA)/gamma-seminoprotein ratio in the cases with PSA levels less than or equal to 10.0ng/ml. Nippon Hinyokika Gakkai Zasshi 86:296–303.
42. Catalona WJ, Smith DS, Wolfert RL, et al. 1995. Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening. JAMA 274:1214–1220.
43. Wun LM, Feuer EJ, Miller BA. 1995. Are increases in mammographic screening still a valid explanation for trends in breast cancer incidence in the United States? Cancer Causes Control 6:135–144.
44. Chu KC, Tarone RE, Wong-Ho C, et al. 1994. Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990. J Natl Cancer Inst 86:997–1006.
45. Potosky AL, Kessler LG, Gridley G, et al. 1990. Rise in prostatic cancer incidence associated with increased use of transurethral resection. J Natl Cancer Inst 82:1624–1628.
46. Levy IG, Gibbons L, Collins JP, et al. 1993. Prostate cancer incidence trends in Canada: rising incidence or increased detection? Can Med Assoc J 149:617–624.
47. Gilliland F, Becker TM, Smith A, Key CR, Samet JM. 1994. Trends in prostate cancer incidence and mortality in New Mexico are consistent with an increase in effective screening. Cancer Epidemiol Biomarkers Prev 3:105–111.
48. Demers RY, Swanson GM, Weiss LK, Kau T. 1994. Increasing incidence of cancer of the prostate. The experience of black and white men in the Detroit Metropolitan Area. Arch Intern Med 154:1211–1216.
49. Ries LAG, Miller BA, Hankey BF, Kosary CL, Harras A, Edwards BK (eds.). 1994. SEER Cancer Statistics Review, 1973–1991: Tables and Graphs. Bethesda: NIH Pub. No. 94-2789, National Cancer Institute.
50. Potosky AL, Miller BA, Albertsen PC, Kramer BS. 1995. The role of increasing detection in the rising incidence of prostate cancer. JAMA 273:548–552.
51. Breslow N, Chan CW, Dhom G, et al. 1977. Latent carcinoma of prostate at autopsy in seven areas. Int J Cancer 20:680–688.
52. Stemmermann GN, Nomura AM, Chyou PH, Yatani R. 1992. A prospective comparison of prostate cancer at autopsy and as a clinical event: the Hawaii Japanese experience. Cancer Epidemiol Biomarkers Prev 1:189–193.
53. Lu-Yao GL, Greenberg ER. 1994. Changes in prostate cancer incidence and treatment in USA. Lancet 343:251–254.
54. Lu-Yao GL, McLerran D, Wasson J, et al. 1993. An assessment of radical prostatectomy: time trends, geographic variations, and outcomes. JAMA 269:2633–2636.
55. Sherman CR, Potosky AL, Weis KA, Ferguson JH. 1992. The Consensus Development Program. Detecting changes in medical practice following a consensus conference on the treatment of prostate cancer. Int J Techol Assess Health Care 8:683–693.
56. Harlan L, Brawley O, Pommerenke F, et al. 1995. Geographic, age, and racial variation in the treatment of local/regional carcinoma of the prostate. J Clin Oncol 13:93–100.
57. Gilliland FD, Welsh DJ, Hoffman RM, Key CR. 1995. Rapid rise and subsequent decline in prostate cancer incidence rates for New Mexico, 1989–1993. Cancer Epidemiol Biomarkers Prev 4:797–800.
58. Epstein JI, Walsh PC, Carcichael M, Brendler CB. 1994. Pathologic and clinical findings to predict tumor extent of nonpalpable (stage T1c) prostate cancer. JAMA 271:368–374.
59. Black W, Welch H. 1993. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. N Engl J Med 328:1237–1243.
60. Gerber GS, Thompson IM, Thisted R, Chodak GW. 1993. Disease-specific survival following routine prostate cancer screening by digital rectal examination. JAMA 269:61–64.
61. Morrison AS. 1991. Intermediate determinants of mortality in the evaluation of screening. Int J Epidemiol 20:642–650.

62. Etzioni RD, Connor RJ, Prorok PC, et al. 1995. Design and analysis of cancer screening trials. Stat Methods Med Res 4:3–17.
63. Frisell J, Eklund G, Hellstrom L, Somell A. 1987. Analysis of interval breast carcinomas in a randomized screening trial in Stockholm. Breast Cancer Res Treatment 9:219–225.
64. Fontana RS. 1986. Screening for lung cancer: recent experience in the United States. In Hansen HH (ed.), Lung Cancer: Basic and Clinical Aspects. Boston: Martinus Nijhoff, pp. 91–111.
65. Johansson JE, Adami HO, Andersson SW, et al. 1992. High 10-year survival rate in patients with early, untreated prostatic cancer. JAMA 267(16):2191–2196.
66. Adolfsson J. 1993. Deferred treatment of low grade stage T3 prostate cancer without distant metastases. J Urol 149(2):326–329.
67. Prorok PC, Connor RJ, Baker SG. 1990. Statistical considerations in cancer screening programs. Urol Clin North Am 17:699–708.
68. Madsen PO, Graversen PH, Gasser TC, Corle DK. 1988. Treatment of localized prostatic cancer. Radical prostatectomy versus placebo. A 15-year follow-up. Scand J Urol Nephrol 110 (Suppl):95–100.
69. Graverson PH, Nielsen KT, Gasser TC, et al. 1990. Radical prostatectomy versus expectant primary treatment in stages I and II prostatic cancer. A fifteen-year follow-up. Urology 36:493–498.
70. Moon T, Brawer M, Wilt T. In press. Prostate Intervention Versus Observation Trial (PIVOT): A randomized trial comparing radical prostatectomy versus palliative expectant management for the treatment of clinically localized prostate cancer. J Natl Cancer Inst (Monogr).
71. Wasson JH, Cushman CC, Bruskewitz RC, et al. 1993. A structured literature review of treatment for localized prostate cancer. Arch Family Med 2:487–493.
72. Mold JW, Holtgrave DR, Bisonni RS, et al. 1992. The evaluation and treatment of men with asymptomatic prostate nodules in primary care: a decision analysis. J Family Pract 34:561–568.
73. Fleming C, Wasson JH, Albertsen PC, et al. 1993. A decision analysis of alternative treatment strategies for clinically localized prostate cancer. JAMA 269:2650–2658.
74. Beck J, Kattan M, Miles B. 1994. A critique of the decision analysis for clinically localized prostate cancer. J Urol 152:1894–1899.
75. Optenberg SA, Wojcik BE, Thompson IM. 1995. Morbidity and mortality following radical prostatectomy: a national analysis of civilian health and medical program of the uniformed services beneficiaries. J Urol 153:1870–1872.
76. Fowler FJ, Roman A, Barry MJ, et al. 1993. Patient-reported complications and follow-up treatment after radical prostatectomy: the National Medicare experience: 1988–1990 (updated June 1993). Urology 42:622–629.
77. Jonler M, Messing EM, Rhodes PR, Bruskewitz RC. 1994. Sequelae of radical prostatectomy. Br J Urol 74:352–358.
78. Litwin MS, Hays RD, Fink A, et al. 1995. Quality-of-life outcomes in men treated for localized prostate cancer. JAMA 273:129–135.
79. Optenberg SA, Thompson IM. 1990. Economics of screening for carcinoma of the prostate. Urol Clin North Am 17:719–737.
80. Littrup PJ, Goodman AC, Mettlin CJ, et al. 1993. The benefit and cost of prostate cancer early detection. CA 43:134–149.
81. Krahn M, Mahoney JE, Eckman M, et al. 1994. PSA screening for prostate cancer: a decision analytic view. JAMA 272:773–780.
82. U.S. Congress, Office of Technology Assessment. 1995. Costs and Effectiveness of Prostate Cancer Screening in Elderly Men, OTA-BP-H-145. Washington, DC: U.S. Government Printing Office, May.
83. Riley GF, Potosky AL, Lubitz JD, Kessler LG. 1995. Medicare payments from diagnosis to death for elderly cancer patients by stage at diagnosis. Med Care 33(8):828–841.

84. Taplin SH, Barlow W, Urban N, et al. 1995. Stage, age, comorbidity and direct costs of colon, prostate, and breast cancer care. J Natl Cancer Inst 87:417–426.
85. Auvinen A, Rietbergen JB, Denis LJ, et al. In press. Prospective evaluation plan for randomized trials of prostate cancer screening. J Med Screening.

8. Screening for gastric cancer

Paola Pisani and D. Maxwell Parkin

1. Introduction

The decline of the incidence of stomach cancer in western countries is a well-known phenomenon [1]. Nevertheless, in 1985, it was still the second most common cancer in the world, and the most common cancer in developing countries [2]. Symptoms are rather non-specific, so most cases are diagnosed when the tumor has already invaded the muscular layer. The prognosis is then very poor: five-year relative survival was 18% for all incident cases in the U.S.A. around 1987 [3], and it was 19% in Europe around 1984 [4]. In contrast, survival is extremely good for early cancers: in the Japanese series, the five-year relative survival is 98% [5], and in the U.S.A. it is 70% [6].

2. Screening for stomach cancer in Japan

The vastly superior survival of individuals diagnosed when their cancers are confined to the gastric mucosa and submucosa was the rationale supporting the introduction of screening for gastric cancer in Japan, where, in the 1970s and early 80s, mortality rates were the highest in the world [7]. The program was introduced in the 1960s and was gradually extended to the whole country; the aim was to examine 30% of the population aged 40 or over each year by photofluorography [8]. The proportion of cancers diagnosed at early stages thereafter increased progressively — from about 15% to over 60% in Miyagi prefecture [9] — proving the capacity of the radiographic test to detect early cancer.

Unfortunately, the program was introduced as a community service before any formal evaluation could provide evidence of the efficacy of the intervention, and improved survival, as frequently mentioned in this book, is a biased measure of efficacy. In fact, there is evidence that malignant transformation of cells of the gastric mucosa and invasion are preceded by detectable precancerous lesions, the evolution of which occurs in a time span of years [10,11]. Improved survival could therefore be the consequence of lead-time bias or length bias for screen-detected cases [12].

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

In common with almost all countries, mortality rates from stomach cancer in Japan have been in decline in most age groups for many years [13]. Nevertheless, when examined for different areas of the country, the rate of decline did appear to be associated with the intensity of screening [14].

Hisamichi et al. [9] have analyzed time trends of incidence and mortality for stomach cancer during the period 1960–1985 in Miyagi prefecture, for which population morbidity and mortality statistics are available. Both measures follow the general trend towards a decline of the disease, but the rate of decline was greater for mortality. This observation has been interpreted as a proof of the success of the mass screening. However, incidence could have been inflated by the screening intervention, which may cause the detection of some cancers that would otherwise not be diagnosed within the lifespan of some individuals.

That overdiagnosis is possible is suggested by the results of a study in Osaka prefecture [15]. Thirty-three thousand individuals screened at least once during a three-year period were followed up by the cancer registry for an average of 6.1 years. The number of gastric cancer cases diagnosed in the cohort was 1.46 times higher than expected, based on the general population rates, while there was a modest reduction in the number of deaths (O:E = 0.91 for all age groups). The results also suggest the presence of selection bias in those attending for screening, since the number of other cancers diagnosed in the cohort was only 0.64 of that expected (a "healthy screenee" effect). A survey in Miyagi prefecture [16] observed that, compared to nonparticipants, those taking part in the gastric cancer screening program had a lower prevalence of smoking and higher intakes of dietary items such as milk and fruit (associated with a reduced risk of gastric cancer).

2.1. Case–control studies of stomach cancer screening in Japan

Selection bias is a well-known source of difficulty in the interpretation of case–control studies also [17,18]. Three such studies have been reported from Japan (table 1). The first study was conducted in Nose Town (Japan), an area covered by the population-based cancer registry of Osaka [19], which allowed the identification of all residents who died from stomach cancer in the period 1969–1981. A total of 87 dead cases and three times as many controls were included in a matched analysis. Controls were live individuals, matched to the cases by sex, age, and residence. Examinations within the 12 months preceding the diagnosis of the case were excluded from both cases' and controls' screening histories, in order to exclude gastric cancer cases who attended because of symptoms. The odds ratio (OR) for screened vs. non-screened subjects was 0.59 (90% confidence interval, 0.34–1.04) in men and 0.38 (0.18–0.78) in women, showing a relative protection of about 50%. The protective effect was greater for those who had more than one examination (the linear trend of the ORs was significant) and seemed somewhat greater also for those who had had a test within two years of the date of diagnosis of the case.

Table 1. Case–control studies evaluating gastric-cancer screening efficacy

Reference and country	Cases	Controls	OR (C.I.)[a]	Exposure definition: tests before case diagnosis
Oshima et al. [19], Japan	87 deaths from gastric cancer	261 live controls matched for sex, age, and residence	0.59 (0.34–1.04) 0.38 (0.18–0.78)	All tests, excluding preceding year: men women
Fukao et al. [20], Japan	367 advanced gastric cancers	367 live controls matched for sex, age, and residence	0.34 (0.25–0.48)	
Fukao et al. [21], Japan	198 deaths from gastric cancer	577 live controls matched for sex, age, and residence	0.41 (0.28–0.61) 0.22 (0.09–0.51)	Tests within preceding 5 years: at least 1 4–5 tests
Pisani et al. [22], Venezuela	241 deaths from gastric cancer	2041 live controls matched for sex, age, and residence	1.26 (0.83–1.91)	All tests, excluding preceding 6 months
	85 ever-screened only	375 ever-screened only	0.25 (0.12–0.51)	

[a] Confidence Interval. All cited are 95%, except for those for the study of Oshima et al. [19], which are 90%.

The same sampling design was adopted in the second study, which compared the screening histories of 367 cases of advanced stomach cancer with the same number of controls matched for sex, age, and residence [20]. A significant protective effect was detected up to three years after the last negative test: OR = 0.34 (95% CI, 0.25–0.48).

A third case–control analysis of the efficacy of the Japanese program has been published recently [21]. The 198 dead cases and 577 live controls matched by sex, age, and residence (district) were compared with respect to their screening histories in the five years preceding the case's diagnosis. The authors considered that this was the relevant exposure period, according to the results of their earlier case–control study. Recent examinations were associated with a relative protection of 59% (OR = 0.41; 95% CI, 0.28–0.61). No significant difference was observed by year since the previous examination within the five years considered, but the relative protection increased to 88% (41%–91%) for those who had had four or five tests compared to none.

2.2. Randomized trials of gastric cancer screening

Only one trial has so far been implemented [9]. In 1985, 39 municipalities in Miyagi prefecture, Japan, were randomized to two groups, A and B. In group A, individuals aged 50–59 were sent an invitation especially written by the head of local government to attend gastric cancer screening, while in group B, individuals aged 60–69 were so invited. The objective was to follow mortality rates in the intervention and control populations. However, results published

to date show that compliance with screening was higher in intervention than control areas only in the first year — especially in the 50–59 age group (about 30% vs. 15%) — and the differential had largely disappeared in subsequent years. It is inconceivable that, with such a small difference in screening experience between the intervention and control groups, any difference in mortality will be demonstrable in this trial. Since gastric cancer screening is considered a routine procedure in Japan, no trial that includes a relatively unscreened control group will be feasible there, and the only possibility is to test the procedures in a high-risk population elsewhere (in east Asia or Latin America).

3. Screening for stomach cancer in other countries

There has been little systematic screening for gastric cancer outside Japan. However, a program using the same methodology was introduced in Tachira state, Venezuela, in 1980 [22]. The radiographic equipment was installed in two mobile units through which screening was potentially offered to the whole population aged 35 and above. The 241 cases who died from stomach cancer in the period 1985–1989 and who could be confirmed by clinical documentation were compared to 2410 controls, drawn from the electoral rolls of the same district of residence as the cases (matched by sex and age). An excess risk of dying from stomach cancer was observed for those whose tests occurred within the six months preceding the diagnosis of the case. No benefit from screening was detected when recent examinations were excluded from both the cases' and the controls' screening histories. An analysis limited to subjects ever screened showed a relative benefit of 53% (OR = 0.47; 95% CI, 0.24–0.98), and most of this effect was confined to tests undergone within the three years preceding the cases' diagnosis. These results suggest that the program was used primarily by symptomatic individuals at high risk of death from gastric cancer, and the resulting selection bias masks any potential benefit. This effect probably results from the low population coverage — only 16% of the control subjects had ever been screened. In contrast, the programs in Japan reach a much larger proportion of the population, e.g., 69% ever-screened in Osaka in 1983 [19] and 78% screened in the preceding five years in Miyagi in the mid-1980s [16].

4. Other approaches to screening for gastric cancer

A different approach to screening, based upon serological markers of lesions of the gastric mucosa that are associated with an increased risk of gastric cancer, has recently been proposed. These "precancerous lesions" are atrophic gastritis, intestinal metaplasia, and dysplasia. The markers have been identified since the 1970s as pepsinogens A (PGA) and pepsinogens C (PGC) [23–

25]. Atrophic gastritis of the fundic gland mucosa is associated with decreased serum levels of PGA and normal or increased levels of PGC. Although the reason for the findings is not well understood, the validation of these markers for the presence of chronic atrophic gastritis has generally given consistent results in population groups showing different background risks for stomach cancer, namely, the U.S., Finland, Netherlands, England, and Japan [26–31]. Nevertheless, there is little agreement on the optimal cut-off points and on the resulting sensitivity and specificity of the tests in identifying more advanced and specific metaplasia, dysplasia, and cancers of the intestinal type [26,31,32].

As a screening test for cancer, pepsinogen levels alone would not be suitable, due to their low sensitivity and specificity. The sensitivity and specificity of the two antigens for gastric cancer of either the intestinal or diffuse type has been measured in a cross-sectional study, which included 19 healthy volunteers and 379 individuals referred for gastroscopy because of symptoms of gastric disorders [28]. Sensitivity could be as high as 83%, but the corresponding specificity was only 19%. Widening the definition of "positives" to include not only gastric cancer but also atrophic gastritis and intestinal metaplasia lowered sensitivity to 65% and improved specificity to 24%. Similarly, in a follow-up of 7488 Japanese-Americans living in Hawaii, Stemmerman et al. [32] found that, depending on the threshold fixed to partition the population, the combinations of sensitivity and specificity of pepsinogen levels for predicting development of invasive gastric cancer within a maximum of 13 years later ranged between 76% and 61%, and 94% and 23%, respectively.

These figures clearly show that the predictive value of serum pepsinogens for gastric tumors is too poor for the purpose of screening; however, they may be suitable for selecting a high-risk population for the radiographic test. In this respect, it would also be worth studying the value of markers of infection with *Helicobacter pylori* in selecting high-risk individuals. Infection with the bacterium was accepted by an IARC Working Group as a cause of gastric cancer [33], and prevalent infection can be detected by serum antibody levels. However, the prevalence of seropositive individuals is high in many populations [34,35], and, although seropositivity predicts some excess risk, it is neither sensitive nor specific enough for screening purposes.

References

1. Howson CP, Hiyama T, Wynder EL. 1986. The decline in gastric cancer: epidemiology of an unplanned triumph. Epidemiol Rev 8:1–27.
2. Parkin DM, Pisani P, Ferlay J. 1993. Estimates of the worldwide incidence of eighteen major cancers in 1985. Int J Cancer 54:594–606.
3. Ries LAG, Miller BA, Hankey BF, Kosary CL, Harras A, Edwards BK (eds.). 1994. SEER Cancer Statistics Review, 1973–1991: Tables and Graphs (NIH Pub. No. 94-2789). Bethesda, MD: National Cancer Institute.
4. Berrino F, Sant M, Verdecchia R, Capocaccia T, Hakulinen T, Estève J. 1995. Survival of Cancer Patients in Europe: The EUROCARE Study (IARC Sci. Publ. No. 132). Lyon: International Agency for Research on Cancer.

5. Japanese Research Society for Gastric Cancer. 1985. The General Rules for the Gastric Cancer Study, 11th edition. Tokyo: Kanehara Shuppan (in Japanese).
6. Bringaze WL, Chappuis CW, Cohn I, Correa P. 1986. Early gastric cancer: 21-year experience. Ann Surg 204:103–107.
7. Kurihara M, Aoki K, Hisamichi S. 1989. Cancer Mortality Statistics in the World 1950–1985. Nagoya: University of Nagoya Press.
8. Oshima A. 1988. Screening for stomach cancer: the Japanese program. In Chamberlain J, Miller AB (eds.), Screening for Gastrointestinal Cancer. Toronto: Hans Huber Publishers (for UICC), pp. 65–70.
9. Hisamichi S, Fukao A, Sugawara N, Nishikouri M, Komatsu S, Tsuji I, Tsubono Y, Takano A. 1991. Evaluation of mass screening programme for stomach cancer in Japan. In Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.), Cancer Screening. Cambridge, U.K.: Cambridge University Press (for UICC), pp. 357–370.
10. Fujita S, Hattori T. 1977. Cell proliferation, differentiation, and migration in the gastric mucosa: a study on the background of carcinogenesis. In Farger E (ed.), Pathophysiology of Carcinogenesis in Digestive Disease Organs. Tokyo: University of Tokyo Press, pp. 21–36.
11. Correa P. 1992. Human gastric carcinogenesis: a multistep and multifactorial process — First American Cancer Society Award lecture on cancer epidemiology and prevention. Cancer Res 52:6735–6740.
12. Morrison AS (ed.). 1985. Screening in Chronic Disease. Oxford: Oxford University Press.
13. Kuroishi T, Hirose K, Tajima K, Tominaga S. 1994. Cancer Mortality in Japan (1950–1990). GANN Monogr Cancer Res 41:1–105.
14. Kuroishi T, Hirose K, Nakagawa N, Tominaga S. 1983. Comparison of time trends of stomach cancer death rates between the model area of the screening program and the control area. J Gastroenterol Mass Survey 58:45–52 (in Japanese).
15. Oshima A, Hanai A, Fujimoto I. 1979. Evaluation of a mass screening program for stomach cancer. NCI Monogr 53:181–186.
16. Fukao A, Hisamichi S, Komatsu S, Shimizu H, Satoh H, Nakatsuka H, Watanabe T, Fujisaku S, Ichinowatari Y, Kuroda S, Ida Y, Suda S, Kato K, Ikeda M. 1992. Comparison of characteristics between frequent participants and non-participants in screening program for stomach cancer. Tohoku J Exp Med 166:459–469.
17. Moss SM. 1991. Case–control studies of screening. Int J Epidemiol 20:1–6.
18. Friedman DR, Dubin N. 1991. Case–control evaluation of breast cancer screening efficacy. Am J Epidemiol 133:974–984.
19. Oshima A, Hirata N, Ubukata T, Umeda K, Fujimoto I. 1986. Evaluation of a mass screening program for stomach cancer with a case–control study design. Int J Cancer 38:829–833.
20. Fukao A, Hisamichi S, Sugawara N. 1987. A case–control study on evaluating the effect of mass screening on decreasing advanced stomach cancer. J Gastroenterol Mass Survey 75:112–116 (in Japanese).
21. Fukao A, Tsubono Y, Tsuji I, Hisamichi S, Sugahara N, Takano A. 1995. The evaluation of screening for gastric cancer in Miyagi prefecture, Japan: a population-based case–control study. Int J Cancer 60:45–48.
22. Pisani P, Oliver WE, Parkin DM, Alvarez N, Vivas J. 1994. Case–control study of gastric cancer screening in Venezuela. Br J Cancer 69:1102–1105.
23. Samloff IM. 1971. Pepsinogens, pepsins and pepsin inhibitors. Gastroenterology 60:586–604.
24. Samloff IM. 1982. Pepsinogens I and II: purification from gastric mucosa and radioimmunoassay in serum. Gastroenterology 82:26–33.
25. Samloff IM, Secrist DM, Passaro E. 1975. A study of the relationship between serum group I pepsinogen levels and gastric acid secretion. Gastroenterology 69:1196–2000.
26. Nomura AMY, Stemmermann GN, Samloff IM. 1980. Serum pepsinogen I as a predictor of stomach cancer. Ann Intern Med 93:537–540.
27. Samloff IM, Varis K, Ibamaki T, Sicerala M, Rotter JI. 1982. Relationships among serum pepsinogen I, serum pepsinogen II and gastric mucosal histology: a study in relatives of patients with pernicious anemia. Gastroenterology 83:204–209.

28. Westerveld BD, Pals G, Lamers CBHW, Defize J, Pronk JC, Frants RR, Ooms ECM, Kreuning J, Kostense PJ, Eriksson AW, Meuwissen SGM. 1987. Clinical significance of pepsinogen A isozymogens, serum pepsinogen A and C levels, and serum gastrin levels. Cancer 59:952–958.
29. Miki K, Ichinose M, Shimizu A, Huang SC, Oka H, Furihata C, Matsushima T, Takahashi K. 1987. Serm pepsinogens as a screening test of extensive chronic gastritis, Gastroenterol Jpn 22:133–141.
30. Miki K, Ichinose M, Kawamura N, Matsushima M, Ahmad HB, Kimura M, Sano J, Tashiro T, Kakei N, Oka H, Furihata C, Takahashi K. 1989. The significance of low serum pepsinogen levels to detect stomach cancer associated with extensive chronic gastritis in Japanese subjects. Jpn J Cancer Res 80:111–114.
31. Knight T, Wyatt J, Newell D, Hengels K, Corlett M, Wilson A, Greaves S, Webb P, Forman D, Elder JB. 1993. Use of serum pepsinogens and *Helicobacter pylori* serology for assessment of antral gastritis in population-based studies of gastric carcinogenesis (abstract). Br Soc Gastroenterol S5.
32. Stemmermann GN, Samloff IM, Normura AMY, Heilbrun LK. 1987. Serum pepsinogens I and II and stomach cancer. Clinica Chimica Acta 163:191–198.
33. IARC. 1994. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 61. Schistosomes, Liver Flukes and *Helicobacter pylori*. Lyon: International Agency for Research on Cancer.
34. Mégraud F, Brassens-Rabbé MP, Denis F, Belbouri A, Hoa DQ. 1989. Seroepidemiology of *Campylobacter pylori* infection in various populations. J Clin Microbiol 27:1870–1873.
35. The Eurogast Study Group. 1993. An international association between *Helicobacter pylori* infection and gastric cancer. Lancet 341:1359–1362.

9. Screening for lung cancer

D. Maxwell Parkin and Paola Pisani

1. Introduction

Lung cancer appears to be an obvious priority for screening programs, according to many of the well-known evaluative criteria [1]. It is currently the most common cancer worldwide, both in terms of annual numbers of cases [2] and deaths [3]. Survival from lung cancer is poor — only about 12% for men in the United States [4], rather lower (8%) in European centers [5], and even worse in developing countries [6]. Several early attempts were made to screen for lung cancer in uncontrolled projects that mainly relied upon regular chest x-rays — for example examinations every six months in the Veterans Administration study [7], the Philadelphia Pulmonary Neoplasm Research Project [8], and the South London Lung Cancer Study [9]. These, and the more recent controlled studies referred to below, confirmed that regular chest x-ray examination could detect tumors that were at an earlier stage (smaller and with less spread), and that these had better survival than cancers detected following clinical presentation.

2. Screening by cytological examination of sputum samples

The clear findings from the early studies seemed to provide (and still provide, for many clinicians) powerful evidence for the effectiveness of screening. However, by the early 1970s, when the National Cancer Institute collaborative trials on lung cancer screening were being initiated, interest had largely moved to the possible benefit that could be achieved by adding cytological screening of sputum to regular chest x-rays [10]. This was a pity, because it is now self-evident that the components of these early results attributable to biases of selection, lead time, length bias, and overdiagnosis [11,12] had scarcely been addressed at all. Although sputum cytology can detect asymptomatic lung cancer — including cases not visible on chest x-ray — screening by cytological examination of sputum samples has been uniformly unsuccessful in reducing lung cancer mortality. Thus, the two trials in the cooperative early lung cancer detection program, in which four-month sputum examinations were added to

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

annual chest x-rays in the intervention group (the Memorial Sloan Kettering Study [13,14], and the John Hopkins Lung Project [15]), did not demonstrate any reduction in mortality in the group receiving cytological examinations. It is possible that immunocytochemical detection of tumor-specific surface antigens on sputum cells will provide a higher specificity (and positive predictive value) than cytology [10], but the value of these newer approaches remains to be demonstrated.

In the remainder of this brief review, we shall examine the current state of evidence for the effectiveness of screening by chest x-ray examinations.

3. Screening by chest x-ray examination

Uncertainty about the efficacy of lung cancer screening has undoubtedly inhibited its introduction as a widespread routine service, so retrospective observational studies are few in number. One case–control study in the former German Democratic Republic [16,17] showed no difference in screening history in lung cancer deaths and in controls, while a second in Japan [18] suggested a small (nonsignificant) reduction in the risk of death (odds ratio (OR), 0.72) in relation to very recent (interval <12 month) examinations. In view of the problems of interpreting case–control studies — in particular, the potential for selection bias [19–21] — it is fortunate that, for lung cancer, five controlled trials have been reported in which individuals have been allocated to intervention (screened) or control groups and self-selection has played no part in the outcome.

These trials are listed in table 1. In none of them was the control group completely unscreened, although in two studies (the North London and Czech trials) the controls must have received very little screening between enrollment (at which an x-ray of all subjects was performed to eliminate "prevalent" cancers) and an end-of-study examination three years later.

The Erfurt county study [22] is the least satisfactory. It compared the male population aged 40–65 living in four counties and offered six-month screenings (41,532 subjects) with a control group (102,348) in the other 10 counties, for whom screening at one- to two-year intervals was offered. Selection of study and control areas was presumably nonrandom, and it is not clear how comparable these areas were, although lung cancer incidence prescreening and mortality from all causes during the trial were similar. Lung cancer mortality can be estimated, and was slightly higher in the intervention group than in the controls during the seven-year study (0.76 per 1000 person-years (py) vs. 0.73 per 1000 py).

The two studies in U.S. compared groups offered regular chest x-ray screening with control populations receiving "regular" care (and a lower frequency of examination).

The Kaiser Permanente study [23] compared 5138 individuals (both sexes) aged 35–54 at entry receiving annual multiphasic health checkups (MHCs),

Table 1. Trials of lung cancer screening using chest x-ray: design and results

Study	Subjects	Number intervention/control	Intervention	Controls	Follow-up	Mortality intervention/control		
						Deaths	Rate (per 1000 py)	R.R.
Erfurt county [22]	Males 40–65 in 14 counties	4 counties (41,532)/ 14 counties (102,348)	6-monthly CXR	CXR every 1–2 yrs	7 yr	209/486	0.76/0.73	1.04
Kaiser Permanente [23]	Both sexes 35–54	5138/5536	Annual CXR [avg = 0.43 pa]	"Usual care" [avg = 0.18 pa]	16 yr	44/42	0.54/0.48	1.13
Mayo Lung Project [25]	Males 45+ Heavy smokers	4618/4593	4-monthly CXR+ sputum cytol. for 6 yrs [compliance = 75%]	Advised annual exams [73% recd. at least one CXR in last 2 years of intervention]	11 yr	122/115	3.2/3.0	1.07
North London study [26]	Males 40+ in 119 factories	75 factories (29,723) 44 factories (25,311)	6-monthly CXR for 3 yrs [compliance = 75%]	Initial prevalence CXR Final CXR at 3 yrs (63% received)	3 yr	62/59	0.70/0.79	0.89
Czech trial [28]	Males 40–64 Heavy smokers	3171/3174	6-monthly CXR for 3 yrs [compliance 90–95%]	Initial prevalence CXR Final CXR at 3 yrs	6 yr	64/47	3.6/2.6	1.38

including a chest x-ray examination, with 5536 controls left to make their own arrangements for checkups. Assignment was random, and, over the next 16 years (1965–1980), screening was considerably more intense in the MHC population (84.3% with at least one examination; average number, 6.8) than the controls (63.8% screened; average, 2.8 times). Despite this, there was no significant difference in mortality rates from lung cancer between the MHC group (0.54 per 1000 py) and controls (0.48 per 1000 py).

The Mayo Lung Project [24,25] began in 1971, and 9211 men aged 45 and over with a history of heavy cigarette smoking were randomized. The intervention group received chest x-rays and sputum cytology every four months for a six-year period, with full compliance of 75%. The control group was advised at enrollment to have yearly chest x-rays and sputum tests — advice that was quite well followed, since 73% of these subjects had at least one chest x-ray in the last two years of the intervention period. Postscreening observation continued for an average of three years, so about 38,000 py of observation accrued in both groups. Neither at the end of the intervention period nor at the end of follow-up was there any difference in mortality between the two groups (3.2 per 1000 py in the intervention group; 3.0 in controls).

The first randomized controlled trial of lung cancer screening [26,27] included 55,034 men aged 40 or more working in 119 factories in north London. Randomization was by factory (stratified by type and locality) and was very successful in producing two groups very similar in age structure and smoking history. Both intervention and control groups received an initial prevalence x-ray, and subjects who were negative were randomized to receive either screening every six months or nothing. A final x-ray was performed on everybody after three years. Compliance seems to have been good, with the intervention group receiving 75% of the scheduled screening x-rays. Of the controls, 63% appeared for their final three-year exam, and it is very unlikely that they received any x-rays, except for diagnostic purposes, outside the trial. Follow-up of the two groups to three years was 99% complete, during which time mortality rates from lung cancer were 0.70 per 1000 person years in the screened group, and 0.79 in the controls (not significantly different).

A more recently reported randomized controlled trial in Czechoslovakia [28] was somewhat similar in design to the North London trials (initial prevalence examination, screening every six months in the intervention group, final exam of both groups at three years). However, the 6364 subjects were confined to male smokers aged 40–64, individual randomization was performed, and follow-up of both groups was continued for three years postintervention with annual examinations. Compliance with screening was high (90% to 95%), and virtually no screening examinations of control subjects took place. At the end of six years, mortality rates from lung cancer were 3.6 per 1000 py in the intervention group and 2.6 per 1000 py in the controls.

Four of these studies (no data are available for the Kaiser Permanente study) confirmed that the lung cancers detected by screening were smaller in size and more usually resectable, and these cases had higher five-year survival

Table 2. Incidence of lung cancer in trials of screening

Trial	Duration of folllow-up	Person-yrs	Cases		OR (95% C.I.)
			No.	Rate/1000py	
Erfurt county [22]	7yr	I: 274,500 C: 670,800	I: 374 C: 667	1.4 1.0	1.37 (1.21, 1.55)
Mayo Lung Project [25]	11yr	I: 37,800 C: 37,700	I: 206 C: 160	5.5 4.3	1.28 (1.04, 1.57)
North London study [26]	3yr	I: 88,248 C: 75,132	I: 101 C: 76	1.1 1.0	1.13 (0.84, 1.52)
Czech trial [28]	6yr	I: 17,880 C: 18,080	I: 108 C: 82	6.0 4.5	1.33 (1.00, 177)

I = Intervention; C = Control.

than cases detected by nonscreening examinations (which were often performed because of symptoms). These differences, of course, appear in any comparison of the intervention and control groups. Of considerable interest, too, was the observation of higher incidence (detection) rates of lung cancer in the intervention populations (table 2). In both the Mayo Lung Project and the Czech trial, this higher incidence appeared throughout the screening period and persisted unchanged after the intervention ceased.

The remarkable consistency of the results of these trials seems to point to a lack of any reduction in mortality as a consequence of screening. The more favorable outcome for the screen-detected cancers is presumably a consequence of lead time and length-biased sampling and, more importantly, of overdiagnosis bias in the screened population [12]. Overdiagnosis occurs in two circumstances. The first is when screening examinations detect lesions that, had they not been found, would never have caused clinical disease in the individual. This form of pseudo-disease could even be associated with an increase in mortality from "lung cancer," if there were any excess mortality associated with treatment. The second type of overdiagnosis results from the detection by screening of cancers in individuals who, if unscreened, would have died of a different cause before the cancer came to light. Because cancer has been diagnosed, it is certified as a cause of death; had it remained undetected, a different cause would be allocated. This is a far from implausible proposition in populations of heavily smoking middle-aged and elderly men with high death rates from cardiovascular, cerebrovascular, and chronic obstructive pulmonary disease. Lung cancers unsuspected during life may be found in a surprisingly high percentage of autopsies of smokers. In Trieste, Italy, where a high percentage of deaths are autopsied, Delendi et al. [29] found that 30% of lung cancers diagnosed at autopsy were found in men who had been certified as dying of a different condition. An autopsy series from Yale [30] found that 31% of lung cancer cases found at autopsy had not been diagnosed during life (and in rather more than half, cancer had not even been

suspected), while in Scotland [31] the percentage was 34%. It is clear that there is quite a pool of undetected lung cancer, and the excesses of 15% to 35% in lung cancer diagnoses in intensely screened populations are broadly what might be expected.

In any case, it is very unlikely that the increased incidence is due to chance, as has been suggested in a recent article [32]. The statistical probability is small in all of the studies, as shown in table 2, and for the increase to have occurred by chance in all four is extremely unlikely. The pooled estimate of the RR from the four studies listed in table 2 is 1.32 (95% confidence interval, 1.20–1.45).

Recently, the National Cancer Institute of the United States has funded a very large-scale screening trial, including annual chest x-ray examination to prevent lung cancer deaths. The rationale has been described by Smart [33], who draws attention to the familiar observations of favorable stage shift and improved survival in screen-detected tumors. Smart also points to the small size of previous studies and the consequent lack of power to detect small mortality reductions. In fact, power calculations for trials which ignore their actual results is a rather disingenuous approach. Thus, given the results observed in the Czech trial, the probability that a true reduction of 10% in mortality had been missed is 0.01. The pooled relative risk of death (intervention/control) for the five studies combined (table 1) is 1.07 (95% CI, 0.95–1.20), so the probability of having missed a 10% mortality reduction is less than 0.05 (table 2).

4. Conclusion

Currently, there is no evidence whatsoever that lung cancer screening is effective. Even if a small reduction in mortality can be demonstrated in the ongoing U.S. trial, questions will remain as to the cost-effectiveness of this approach. In the future, the possibility of identifying genuinely high-risk groups through genetic markers of susceptibility may mean that screening could be better targeted.

It would, however, be a pity to divert attention from the need to pursue a vigorous approach to primary prevention through curbing tobacco use.

References

1. Wilson JM, Junger G. 1968. Principles and Practice of Screening for Disease. Public Health Papers No. 34. Geneva: WHO.
2. Parkin DM, Pisani P, Ferlay J. 1993. Estimates of the worldwide incidence of eighteen major cancers in 1985. Int J Cancer 54:594–606.
3. Pisani P, Parkin DM, Ferlay J. 1993. Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. Int J Cancer 55:891–903.

4. Ries LAG, Miller BA, Hankey BF, Kosary CL, Harras A, Edwards BK (eds.). 1994. SEER Cancer Statistics Review, 1973–1991. NIH Publication No. 94-2789. Bethesda, MD: National Cancer Institute.
5. Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T, Estève J. 1995. Survival of Cancer Patients in Europe: The EUROCARE Study (IARC Sci. Publ. No. 132). Lyon: International Agency for Research on Cancer.
6. Sriamporn S, Black RJ, Sankaranarayanan R, Kamsa-ad S, Parkin DM, Vatanasapt V. 1995. Cancer survival in Khon Kaen Province, Thailand. Int J Cancer 61:296–300.
7. Lilienfeld A, Archer PG, Burnett CH, Chamberlain EW, Chazin BJ, Davies D, Davis RL, Haber PA, Hodges FJ, Koprowska I, Kordan B, Lane JT, Lawton AH, Lee L, MacCallum DB, McDonald JR, Milder JW, Naylor B, Papanicolaou GN, Slutzker B, Smith RT, Swepston ER, Umiker WO. 1966. An evaluation of radiologic and cytologic screening for the early detection of lung cancer: A Cooperative Pilot Study of the American Cancer Society and the Veterans Administration. Cancer Res 26:2083–2121.
8. Boucot KR, Weiss KR. 1973. Is curable lung cancer detected by semi-annual screening? JAMA 22:1361–1365.
9. Nash FA, Morgan JM, Tomkins JG. 1968. South London lung cancer study. Br Med J ii:715–721.
10. Mulshine JL, Tockman MS, Smart CR. 1989. Considerations in the development of lung cancer screening tools. J Natl Cancer Inst 81:900–906.
11. Cole P, Morrison AS. 1978. Basic issues in cancer screening. In Miller AB (ed.), Screening in Cancer, UICC Technical Report Series, vol. 40. Geneva: International Union Against Cancer, pp. 7–39.
12. Morrison AS. 1985. Screening in chronic disease. New York, Oxford: Oxford University Press.
13. Flehinger BJ, Melamed MR, Zaman MB, Heelan RT, Kimmel M, Perchick WB, Martini N. 1984. Screening for early detection of lung cancer in New York. In Prorok PC, Miller AB (eds.), Screening for Cancer, UICC Technical Report Series, vol. 78. Geneva: International Union Against Cancer, pp. 123–135.
14. Melamed MR, Flehinger BJ, Zaman MB, Heelan RT, Perchick WA, Martini N. 1984. Screening for early lung cancer. Results of the Memorial Sloan-Kettering Study in New York. Chest 86:44–53.
15. Tockman MS, Levin LM, Frost JK, Ball WC, Stitik FP, Marsh BR. 1985. Screening and detection of lung cancer. In Aisner J. (ed.), Lung Cancer: Contemporary Issues in Clinical Oncology, vol. 3. New York: Churchill Livingstone, pp. 25–40.
16. Ebeling K, Nischan P. 1987. Screening for lung cancer — results from a case–control study. Int J Cancer 40:141–144.
17. Berndt R, Nischan P, Ebeling K. 1990. Screening for lung cancer in the middle-aged. Int J Cancer 45:229–230.
18. Sobue T, Suzuki T, Naruke T, and the Japanese Lung-Cancer-Screening Research Group. 1992. A case–control study for evaluating lung-cancer screening in Japan. Int J Cancer 50:230–237.
19. Connor RJ, Prorok PC, Weeds DC. 1991. The case–control design and the measurement of the efficacy of cancer screening. J Clin Epidemiol 44:1215–1221.
20. Moss SM. 1991. Case–control studies of screening. Int J Epidemiol 20:1–6.
21. Weiss NS, McKnight B, Stevens NG. 1992. Approaches to the analysis of case–control studies of the efficacy of screening for cancer. Am J Epidemiol 135:817–823.
22. Wilde J. 1989. A 10-year follow-up of semi-annual screening for early detection of lung cancer in the Erfurt County, GDR. Eur Respir J 2:656–662.
23. Friedman GD, Collen MF, Fireman BH. 1986. Multiphasic health checkup evaluation: a 16-year follow-up. J Chron Dis 39:453–463.
24. Fontana RS. 1984. Early detection of lung cancer; the Mayo project. In Prorok PC, Miller AB (eds.), Screening for Cancer. UICC Technical Report Series, vol. 78. Geneva: International Union Against Cancer, pp. 107–122.
25. Fontana RS, Sanderson DR, Woolner LB, Taylor WF, Miller WE, Muhm JR, Bernatz PE,

Payne WS, Pairolero PC, Bergstralh EJ. 1991. Screening for Lung Cancer. A critique of the Mayo Lung Project. Cancer 67:1155–1164.
26. Brett GZ. 1968. The value of lung cancer detection by six-monthly chest radiographs. Thorax 23:414–420.
27. Brett GZ. 1996. Earlier diagnosis and survival in lung cancer. Br Med J 4:260–262.
28. Kubik A, Parkin DM, Khlat M, Erban J, Polak J, Adamec M. 1990. Lack of benefit from semi-annual screening for cancer of the lung: follow-up report of a randomized controlled trial on a population of high-risk males in Czechoslovakia. Int J Cancer 45:26–33.
29. Delendi M, Riboli E, Peruzzo P, Stanta G, Cocchi A, Gardiman D, Sasco AJ, Giarelli L. 1991. Comparison of diagnoses of cancers of the respiratory system on death certificates and at autopsy. In Riboli E, Dedendi M (eds.), Autopsy in Epidemiology and Medical Research (IARC Sci. Publ. No. 112). Lyon: International Agency for Research on Cancer, pp. 55–62.
30. McFarlane MJ, Feinstein AR, Wells CK, Chan CK. 1987. The "Epidemiologic Necropsy". JAMA 258:331–338.
31. Cameron HM. 1991. The role of the autopsy in assessing clinical diagnosis. In Riboli E, Delendi M (eds.), Autopsy in Epidemiology and Medical Research (IARC Sci. Publ. No. 112). Lyon: International Agency for Research on Cancer, pp. 75–79.
32. Strauss GM, Gleason RE, Sugarbaker DJ. 1993. Screening for lung cancer re-examined. A reinterpretation of the Mayo Lung Project. Randomized Trial on Lung Cancer Screening. Chest 103:337S–345S.
33. Smart CR. 1993. Annual screening using chest X-ray examination for the diagnosis of lung cancer. Cancer 72:2295–2298.

10. Screening for melanoma

J. Mark Elwood

1. Introduction

Screening for cutaneous melanoma was reviewed by the International Union Against Cancer in 1990, with the conclusion that the *potential* for benefit was considerable, but that the effectiveness for screening was still to be demonstrated [1–3]. In this chapter, selected studies published since then will be reviewed, in the context of trends in melanoma, recommendations for screening, current use of screening, data on the performance of the screening test, and requirements for further research. A previous, detailed review of this topic concluded that screening for melanoma had potential but not demonstrated benefit, had substantial costs and potential hazards, and required further assessment, ideally by a randomized trial [4].

2. Trends in melanoma

In most countries, incidence and mortality from melanoma have been rising for several decades up to recent times [5]. This long-term trend has been accompanied by especially strong increases in the frequency of melanoma on intermittently exposed sites such as the trunk. Recent studies in general show a leveling off or even a fall in mortality rates, particularly at young ages or in recent birth cohorts, while most reports show a continuation of increasing rates of incidence. A reduction in mortality rates in birth cohorts of women born since the early 1930s and of men born since the 1950s has been shown in United States data [6]. By age, the reductions were seen at ages under 40, with a continued increase at ages over 50, and projections suggest that the long-term upward trend in mortality rates should change by the second decade of the twenty-first century [7]. Connecticut incidence data [8] show that the rate of increase of incidence of melanoma slowed substantially in males up to 1989, but not in females; further increases in incidence are projected through the 1990s, with a greater increase in females. In Canada, the incidence rate doubled in women from 1969 and 1988, and increased even more in men. The rise in mortality in women lasted to the mid-1980s; since then it has shown

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

some fall, but in men the upward trend has continued [9]. In Queensland, Australia, incidence rates of invasive melanoma rose by more than 50% in women, and more than doubled in men, between 1980 and 1987 [10]. Over the whole country, incidence increased [11], and the increase in mortality, while much less than that of incidence, averaged 2.5% per year in men and 1.1% per year in women from 1969–1989. A leveling off or even reduction in incidence or mortality rates in recent birth cohorts or in younger adults has also been reported from several other countries. A detailed analysis in Sweden [12] shows stabilization of the previously increasing mortality rates in the 1980s, with this being explicable by improvements in survival rates counteracting the increase in incidence [13]. The Swedish study showed no evidence of drift in diagnostic criteria, but large increases in the proportion of thin melanomas, which they attributed to earlier diagnosis. However, there was also an increase in the proportion of superficial spreading melanoma (SSM) and a decrease in the proportion of acral lentiginous melanoma (ALM), suggesting "changes in unmeasured biological features." Thus these and other recent results [14,15] show that in many countries melanoma incidence and mortality may be beginning to level off, at least in some age groups, but melanoma remains and is likely to remain a serious disease and a major contributor to cancer incidence.

There have been large shifts towards the diagnosis of thinner lesions and improvements in postdiagnosis survival. In some countries, such as Scotland [16], this shift to earlier diagnosis may be due to specific educational programs to increase public and professional awareness, but in other countries there is little evidence of systematic educational programs, and the shift must be occurring by a less systematic increased awareness and improved care of early lesions. In no country has systematic screening been undertaken on a large scale in a way that would contribute substantially to these improvements in survival.

3. Natural history of thin melanoma

Consideration of these incidence and survival trends has suggested that an increasing proportion of lesions removed and classified as early invasive melanoma may have a benign natural history [17], and this possibility has been examined in important studies by Burton and colleagues [18,19]. They have shown, using primarily Australian data with confirmation in New Zealand and Scotland data, that the very rapid recent increases in melanoma incidence have been due primarily to the excision and recording of very thin lesions, and the details of these trends accompanied by a continuing increase in the incidence of thicker lesions suggest that advancement of diagnosis is not the total explanation. These increases have been extreme, and noted in the mid-1980s or later; for example, the incidence of invasive melanoma in one area of New South Wales almost doubled between 1986 and 1988 [19], due primarily to an increase in thin melanomas, and similar results have been seen in several other

Table 1. Recent increases in diagnosis of melanoma in two areas of New Zealand

A. Tauranga

	1980–84	1985–89	% increase
Depth of lesion			
0–0.75 mm	25	92	268
0.76–1.49 mm	28	53	89
>1.5 mm	43	61	42
Total invasive	96	206	115

B. Auckland

	1985	1990	% increase
Depth of lesion			
In situ	71	335	372
0–0.75 mm	114	307	169
0.76–1.49 mm	59	86	46
1.5–3 mm	46	52	13
>3 mm	21	22	5
Total invasive	240	467	95

Note: Data are numbers of cases; changes in population at risk are minimal.

From: Brown and Palmer [20], Elwood and Glasgow [21], and Mullin and Crombie, 1991 (unpublished).

areas. Recent New Zealand data are shown in table 1 [20,21]. Burton et al. [18] suggest that there has been a real increase in the incidence and increasing diagnosis of "a pre-existing, non-metastasizing form of thin melanoma." This has major implications for screening, since it raises the possibility that early diagnosis, whether occurring through general increased professional awareness or due to systematic screening programs, may lead particularly to the removal lesions that appear clinically and pathologically malignant, but which would demonstrate a benign natural history if left in place. This situation should be anticipated with any new screening test for early disease [22], and there are analogies to possibly nonprogressive lesions in the uterine cervix, the breast, and other sites identified by early diagnostic tests. This situation emphasizes the need for careful evaluation of screening in terms of ultimate benefit, judged by the incidence of deeply invasive or metastatic disease and in due course the effect on mortality rates.

4. Recommendations of influential groups about melanoma screening

Screening is being recommended and practiced at present, but with wide diversity of opinions of its value. In the United States, both regular physician skin examination and self-screening are advocated by the American Cancer

Society [23] and the American Academy of Dermatology [24]. Friedman et al. [25], writing in conjunction with these two American organizations, conclude that "a complete annual examination of the skin by a physician is recommended for everyone, supplemented by monthly self-examinations by the patient. Patients with a personal or family history of malignant melanoma, as well as those with dysplastic nevi or any of the other risk factors, should have more frequent examinations by both their physicians and themselves."

In contrast, general-population skin screening is not recommended by the Canadian Task Force on the Periodic Health Examination or by the U.S. Preventive Services Task Force [26,27]. Both groups base their recommendations on a critical review of published evidence. Both, however, recommend skin examinations for high-risk groups; the U.S. group recommends screening by complete skin examination for "persons with a family or personal history of skin cancer, clinical evidence of precursor lesions (e.g. dysplastic nevi, certain congenital nevi), and those with increased occupational or recreational exposure to sunlight." These categories, especially the last, could include a substantial proportion of the population. General population screening is not recommended by the Australian Cancer Society [28] or by the International Union Against Cancer, which had recently published a Melanoma Control Manual [29]. The U.S. National Cancer Institute (NCI) state-of-the-art statement on screening for skin cancer in September 1993 [30] shows that only poor-quality evidence is available. While the NCI report concludes that "evidence from non-experimental studies suggests a decrease in mortality from routine examination of the skin," the evidence quoted refers to the effects of public education programs, rather than screening programs.

In 1992, the National Institutes of Health (NIH) Consensus Development Conference on early melanoma [31] concluded that "there is sufficient evidence to warrant screening programmes for melanoma in the United States." However, the justification for this conclusion is not made clear: indeed, there is no evidence given for the effectiveness, as distinct from the potential effectiveness, of screening. The only evidence quoted is "preliminary evidence from uncontrolled trials in countries such as Scotland"; however, the studies in Scotland do not assess screening, but are studies of educational programs. The NIH group concluded that "there is a reliable screening test" and other conditions for potential benefit, but also note that a randomized trial has not been done, and that "the primary care medical community is not yet adequately prepared for undertaking or responding to patient-screening programmes."

5. Current use of screening

In a detailed population-based survey, in New South Wales, Australia, 17% of 1344 randomly selected subjects had had a skin examination from their family practitioner in the previous year, and 48% reported having used self-screening

[32]. This is a high-prevalence figure, but there is no information on the thoroughness or regularity of screening. A survey in the Gold Coast area, an affluent coastal area of Queensland with extremely high melanoma rates, showed that 76% of 995 randomly selected adults engaged in some form of skin examination, 60% using self-examination and 55% examination by a doctor. About half of these medical consultations were for a general skin examination, and others for consultation regarding a specific lesion. Much of the self-checking behavior was restricted to the more accessible body sites [33]. Only 15% of respondents had actually been taught how to check their skin. In a New Zealand study of 1000 adults in 1992, 65% of respondents had checked their skin for changes in the last year, a third by an examination by a doctor and the rest by self-checks or checks by family; and in a separate New Zealand study of 21-year-olds, 53% reported deliberately checking their own skin for suspicious changes in the last year (Douglass, personal communication).

6. Potential benefits and hazards of screening

The potential benefit of screening is based primarily on the large variation in survival rate by depth of invasion for localized tumors, with a range from 47% to 96% five-year survival relating to depths from 4 mm down to less than 0.76 mm [34]. In high-incidence countries, most melanomas are thin; 52% of melanomas were less than 0.76 mm in Australia in 1989 [35]. This variation in survival rates is very large and is too great to be explained by a lead-time effect. However, the previously noted question of thin melanomas including a proportion that may not be biologically aggressive raises further questions about the validity of the argument that this variation demonstrates the value of early diagnosis.

The potential hazards of screening, such as the effects on health service demands and costs, are also very considerable. Public education campaigns result in a doubling or greater increase in the number of hospital referrals [4]. Open-access skin-examination clinics may result in very high proportions of patients being referred, such as 31% in Massachusetts [36]; however, this is not always the case, as shown by the much lower referral rate of 10% seen in the Netherlands [37]. The proportions of subjects referred for suspected melanoma are much lower than these, but this example illustrates the great difficulty in deciding on what range of conditions detected at a screening examination require referral.

A major issue in programs of screening and also in any trials is therefore the provision of adequate, prompt, and efficient referral services for those who are regarded as abnormal in the screening test, and efforts to ensure a high compliance with the follow-up offered. Studies in Sweden [38,39] found no major psychological problems due to involvement in public melanoma screening or due to assessment as a family member of a melanoma patient.

7. Performance criteria for the screening test: skin examination by a physician

A major difficulty is that the performance characteristics of the proposed screening examinations, whether skin cancer examination by a health professional or self-assessment, are still largely unknown, particularly in regard to the use of these tests in an asymptomatic general population [4,40]. Some studies [41] show high values of sensitivity, specificity, and predictive value, but are based on hospital examinations by dermatologists of patients who have already been selected for referral by primary care physicians.

While many open-access skin clinics provide an assessment only of the lesion presented by the patient, or a limited examination of easily accessible body sites, many lesions will only be found by a more thorough examination [42]. The patient's willingness to undertake a full-body skin examination can be helped by written explanations of the need [43].

More information is needed on the effects of different criteria for referral or biopsy. The clinically used criteria, such as the A,B,C,D,E list (Asymmetry, Border irregular, Color mixed, Diameter greater that 6 mm, Elevation) [44] and the Glasgow seven-point check list [45], are put forward as clinical guides. The decision points using the ABCDE list have not been specified; for the Glasgow list, the advice is that all patients with one or more of the major signs of change in size, shape, or color should be considered for referral, and the presence of minor signs of inflammation, crusting or bleeding, sensory change, or diameter over 7 mm should also encourage referral. Several clinical studies have assessed these criteria as they apply to patients seen at referral centers [4], but their use in a systematic fashion in a screening situation has not been assessed. Since these criteria are based mainly on the experience of specialists, seeing patients already referred from primary care, the criteria do not adequately describe the very early lesions that are most relevant to screening [46]. There are few studies also on interobserver consistency; one small study has shown reasonable agreement in a screening situation for skin lesions such as atypical nevi, but did not include melanomas [47], and another study of pigmented lesions and nevi showed that dermatologists and plastic surgeons showed higher diagnostic accuracy than other physicians [48].

The other screening test proposed is self-screening, and again there is a need for further evaluation of its performance characteristics. In relation to self-examination for nevi, Gruber et al. [49] assessed the ability of dermatology patients without melanoma to record the number of freckles and palpable nevi on their right forearm, and nevi greater that 5 mm in diameter on their entire body. Compared to counts by a dermatologist, specificity ranged from 83% to more than 95% for these three cutaneous markers. Similarly, Lawson et al. [50] showed good correlations between self-counting of large nevi by patients visiting the melanoma screening clinic at the Lawrence Livermore Laboratory and subsequent counts by a physician. These studies suggest that self-examination for skin pigmentary characteristics may be useful in identify-

ing individuals at high risk for melanoma. In contrast, in Australia, a study of 132 subjects from the general population showed a poor correlation between nevus counts on the arm carried out by lay interviewers and by dermatologists [51].

8. Frequency of abnormalities in the general population

An important issue is the prevalence in a general population of the early signs of melanoma given by various authorities. In a large study in New South Wales [32,52], 12% of the 1344 respondents had noticed changes of the ABCDE type in the previous year; 6.8% had sought medical advice, 3.2% had a biopsy, and 0.5% had a melanoma diagnosed. The predictive value of having any of these signs of melanoma was therefore 4.2%. In a study of some 900 subjects aged 21 years in New Zealand, 205 had noticed changes in a mole or freckle in the last 12 months, and 46% of those who did had sought medical advice (Douglass, personal communication). Given that in this age group the five-year cumulative risk of melanoma would only be about one in 1000, the predictive value of these changes is extremely low. Thus a major problem in skin cancer screening is that features that have been used primarily for the clinical assessment of lesions for which the patient has taken the initiative of seeking help, or lesions that have already been seen by a primary care physician and then referred to a specialist, cannot be used uncritically to set up appropriate decision rules for a general-population screening program, where the frequency of disease will be so much lower.

9. Evaluation of screening programs and related activities

9.1. Evidence for the impact of professional and public education programs

MacKie et al. [16,53] have shown a reduction in the incidence and in the mortality of deep melanoma in women in Scotland, which followed intensive public and professional education programs, with no effect on the rates in men. Other less detailed studies from other parts of the United Kingdom, Europe, and Australia also show short-term indicators of the impact of education [4,54]. Such data, which are relevant to the assessment of the effects of public education programs, should not be interpreted as evidence for the benefits of screening.

9.2. Evaluation of open-access skin-check programs

The use of open-access skin-check programs, held at community centers, at locations such as beaches, or in ordinary health care facilities, has been a major component of efforts for earlier diagnosis of skin cancer in many countries,

Table 2. Results of skin examinations performed on population groups and in open access skin screening situations

	Total	Referred (%)	Follow-up	Melanoma confirmed	Predictive value (%)
Netherlands: open access screening by dermatologists, 1989–1990 [37]					
Total number examined	2564				
Referred: total	262	10.2	221	9	4.1
susp. skin cancer	103	4.0	93	8	8.6
other	159	6.2	128	1	0.8
Massachusetts: open-access screening by dermatologists, 1986-1987 [36][a]					
Total number examined	2560				
Referred: total	787	30.7	288	9	3.1
melanoma	26	1.0	22	9	40.9
dysplastic or congenital nevus	197	7.7	110	0	0.0
nonmelanoma skin cancer	236	9.2	156	0	0.0
other	328	12.8	0		
Geraldton, Australia: research survey by dermatologists, 1987 [58]					
Total number examined	4103				
Referred: total	920	22.4	701	20	2.9
melanoma	39	1.0	36	12	33.3
suspicious pigmented lesion	73	1.8	68	2	2.9
nonmelanoma skin cancer	749	18.3	558	6	1.1
other	59	1.4	39	0	0.0

[a] Assumes all 9 melanomas confirmed were in clinical "melanoma" category. Follow-up information was not sought for the "other" category.

and has been reviewed previously [4]. The contrast between high-referral rates in the United States experience and the lower referral rate in the Netherlands has already been noted and is shown in table 2. This later rate probably relates to a different attitude toward screening and toward medico-legal issues. In Europe, Rampen and his colleagues have been able to limit the referrals to lesions suspicious of melanoma, and have not referred so many patients with other suspect lesions, whereas in the United States, the examining dermatologists refer patients with a range of suspected lesions, including, in the earliest report, senile keratoses. The results of these programs in terms of sensitivity depend on how one classifies patients referred for different reasons [4]. This is one of the major limitations of skin checks, and a major potential problem for skin screening programs. In subsequent years, the criteria for referral in the ADD programs have been changed, and results from an extensive follow-up program are being prepared (Koh, personal communication). The other major limitation of these studies as a guide to the effectiveness of the screening is that the patients are self-referred, and in the United States and New Zealand experience, many of those coming have suspicious signs and symptoms [55,56]. The situation may be different in Europe, where Rampen et al. [57] report that those coming to screening clinics were not at higher risk than the general population. The self-selection of subjects will have the effect of increasing the yield of the program and making its operational performance statistics more impressive, but self-selection will limit the impact of the program on a popu-

lation basis, and also limit the applicability of the results to a general population.

9.3. Evaluation of skin examinations of a defined population

There are very few studies that report results from screening campaigns in a defined population. The most detailed is a research study in Geraldton, Western Australia [58], in which 4103 Caucasian subjects — about 76% of all adults aged 40–64 on the electoral roll — were examined by one of nine dermatologists (excluding underwear and hair areas). The results (table 2) showed a referral recommendation of 22% of those examined, although most of these were for basal and squamous cell cancer, with 1% clinically diagnosed as melanoma and 1.8% as another suspicious pigmented lesion. Twenty melanomas were confirmed, of which 12 were clinically diagnosed, and two were in the suspicious pigmented lesion category; the other six were amongst the 514 subjects clinically diagnosed as basal cell cancers. Although these data are useful, since they refer to an unselected population, the thoroughness of the examination and the time spent on it for this detailed research study is likely to have been considerably greater than would be possible in a mass screening situation.

The most comprehensive and thorough education and early diagnosis program is that carried out at the Lawrence Livermore National Laboratory in the U.S.A. over the last several years. This program was set up in response to a suspected increased incidence of the disease. Since 1974, when high awareness of melanoma risk within the work force was accompanied by freely available on-site dermatological examinations, the five-year survival following diagnosis of all melanomas has been 100%. Thus, within the limitations of the size of the group studied and the length of follow-up, complete avoidance of melanoma death has been achieved [59]. The costs of this program are, however, around $150,000 per year for a community of 10,000 subjects, and so could not easily be replicated on a large scale [60].

9.4. Selective screening of high-risk groups

The issue of restricting screening to higher-risk subjects has received much attention, with the main approaches being selection of high-risk subjects on the basis of family history and dysplastic nevi, and secondly selection by other phenotypic characteristics such as total nevus count [4]. The NIH studies have shown very high absolute risks of melanoma in subjects with strong family histories (at least two family members affected) of melanoma, and with the presence of dysplastic nevi, with an eight-year cumulative risk of 7%, although with wide confidence intervals [61]. Intensive surveillance of such subjects by regular clinical examinations aided by photography to monitor changes has shown that further melanomas are diagnosed while they are particularly thin [62–64]. However, the selection and surveillance of these particularly high-risk

subjects, with dysplastic nevi and a family history, will not have a major impact on the whole population problem of melanoma [4,65]. The clinical and pathological definition of atypical nevi or dysplastic nevi is still an area of difficulty, and was the topic of a NIH consensus conference in 1992 [31]. Studies comparing pathologists' assessments of dysplastic nevi show considerable variations [66,67].

The other major method suggested is the selection of subjects at higher risk on the basis of other risk markers. In Australia, having five or more raised moles on the arm is a strong risk indicator, and can be combined with other variables to select high-risk subjects [68]. A similar approach has been suggested in Canada [69]. Such systems may identify between 5% and 15% of a general population, which should contain between 25% and 55% of melanoma subjects [4]. Such methods will increase the yield of selective screening but will reduce its sensitivity considerably, since many or most melanomas will occur in the unscreened lower-risk group. In Scotland, a risk chart has been produced from a case–control study that shows considerable distinctions in risk, but does not allow the estimation of the effect of such selection in population terms [70].

10. The evaluation of melanoma screening

The major issues involved in the assessment of melanoma screening have been reviewed in detail [4]. The screening modality of most interest is skin examination by a health professional, since this is what is being widely recommended. The American experience has emphasized screening by dermatologists, but even in the U.S. it would be logistically impossible for screening of a general population to be carried out by dermatologists. In countries with a stronger primary care sector, screening by general practitioners/family doctors would be possible. A major question is whether the screening involves an invitation to the patient and a visit specifically for that purpose, or is combined with a visit for other purposes. While the former situation is much more amenable to study in a trial and allows for stringent quality control, the latter may be more practical and less expensive. Screening by other health professionals, or even by specifically trained nonprofessionals, is another possibility that has not been explored. The second major modality is self-screening, which is also being recommended by some authorities, but on which there is even less information available.

Any trial needs to pay great attention to the scientific, economic, and ethical issues involved in the decision criteria and the pathways of referral. One solution to the problem of referral would be to have two pathways of referral. Subjects with lesions suspected of being melanoma could be offered immediate expert follow-up by further assessment and biopsy, within the screening trial itself. Those with other lesions found, whose earlier diagnosis is not within the objectives of the trial — such as suspected nonmelanoma skin cancers and

benign lesions — could be advised to go to their normal medical caregiver for further action. Follow-up to check whether this action is taken would be advisable. While the ultimate endpoint of such a trial should be mortality from melanoma, interim endpoints such as the incidence of deep or metastatic melanoma will be useful, and the total incidence of melanoma is also relevant since, as noted above, screening may lead to the detection of apparently malignant but nonprogressive lesions.

10.1. Randomized trials of melanoma screening

The previous review [4] illustrates designs for individually randomized trials and community randomized trials of melanoma screening. A trial using individual randomization, offering screening to 260,000 subjects in two annual rounds with three years of follow-up, is estimated to cost some $10 million, of which 80% is the cost of screening as distinct from the cost of evaluation. A major issue in trial design is therefore the cost of the screening program itself. It is to be hoped that forward-thinking health authorities, who in the absence of a trial may be obligated to spend large sums on screening programs on the basis of current recommendations, will cooperate with researchers by covering the costs of demonstration screening programs, leaving research budgets to contribute the costs of the trial component itself. In practice, it is likely that considerably longer follow-up will be necessary to demonstrate an effect, and larger sample sizes and a longer screening period may be necessary.

The major options for setting up programs that could be evaluated would be to offer screening on a systematic basis to defined groups in a fashion analogous to mammography or cervical cancer screening, through a special facility and program; the difficulty of this is that the screening service needs to be set up alongside and in fact in competition with usual medical care. A second option is to promote screening through family doctors and to design a study comparing individual doctors' practices or communities in which screening is actively promoted with a comparison group. The success of such a trial will depend on the success of improving the uptake of screening dramatically by such promotion. A proposal has been developed in Australia for a trial in which small communities would be selected, screening by family doctors and self-screening would be actively promoted by individual and community approaches, and the incidence and mortality of melanoma in that community would be compared to similar communities without active promotion. The proposal calls for populations of approximately 250,000 in each group in a high-incidence area, and would require at least 10 years of follow-up to demonstrate a difference. The success of the trial in assessing screening would depend on the community-based promotion methods being able to achieve a very major increase in the use of screening compared to the comparison groups, which may be a difficult task in itself.

Another major trial design would be to assess screening in higher-risk subjects, either self-selected or selected by some systematic prescreening pro-

cedure. Such a trial, of course, answers a fundamentally different question that the first option, since the effectiveness of screening in high-risk individuals would not necessarily demonstrate that such screening would be effective in a general community. The results of such a trial would depend considerably on the actual risk level of those screened, and so might be difficult to generalize because the methods and the effects of self-selection into the high-risk group could vary considerably between communities. On the other hand, trial designs that employ even a very simple questionnaire to a general population to allow the selection of those with preset indicators of high risk involve a very substantial cost component, which makes the logistics not greatly different from that of a trial based on a general population.

As noted already, some authorities, particularly in the United States and among dermatologists, take the view that clinical examination is a well-established, simple, low-cost procedure. They also emphasize the difference in postdiagnosis survival by depth of invasion in melanoma, and on this basis argue that skin examination can achieve earlier diagnoses and that such earlier diagnoses must result in reduced mortality and morbidity. They therefore actively recommend screening, primarily by dermatologists, to the public. Moreover, some of these authorities take the view that systematic evaluation of screening, certainly by ambitious and expensive projects such as a randomized trial, is unnecessary. In contrast, other authorities conclude that although skin examination is an accepted and apparently simple technique, its performance is largely unknown, particularly as applied to unselected asymptomatic subjects, and that the technique is not free of cost and risk. In particular, regular skin examinations may lead to high rates of referrals, and the impact of these referrals and their subsequent cost and morbidity needs to be considered. Moreover, while it is accepted that the variation in postdiagnosis survival is substantial, the evidence suggesting that some or perhaps many of these early-diagnosed lesions might not be true precursors of severe and fatal conditions raises further questions about validity. Those who take this view advocate further evaluation, and would argue that the logistic issues and the costs of a large-scale randomized trial, while very substantial, need to be compared with the very large continuing investment that is likely to be made if screening is advocated and performed in the absence of such evaluation.

The difficulties of assessing screening for melanoma are compounded by the simplicity and informality of the screening test. In comparison to mammography or even uterine cervical screening, for example, it is much more difficult to know if an individual has been screened and to assess the timing and quality of that screening. Distinguishing screening from diagnostic examinations is difficult. A reasonable analogy is to breast self-examination, which has also been particularly difficult to assess.

It is therefore the view of this writer that further evaluation of screening for melanoma is very important, and that the continued growing acceptance of skin cancer screening in the absence of such evaluation is likely to lead to much unnecessary health care expenditure and some morbidity, with perhaps

no ultimate benefit. Despite the problems, efforts should be made to set up evaluative studies. Such studies would ideally involve a randomized comparison, probably on a community rather than an individual basis, and might require an international coordinated effort and the active collaboration of health care providers with researchers. While some others agree with this viewpoint [71], there is considerable resistance to such proposals and to the commitment of resources they require [72]. The resistance comes from those who believe that screening is already adequately established, but also from those who feel that the short-term operational characteristics of screening tests have been insufficiently evaluated to mount a trial at this point in time. A major difficulty is that to delay a trial until much better information is available so as to satisfy the second set of critics will make satisfying the first set more difficult, since the more screening is recommended and practiced, the more difficult its evaluation will be.

10.2. Evaluation other than by randomized trails

Apart from randomized trials, other methods of assessing the effectiveness of cancer screening include the comparisons of populations that differ in terms of screening levels. This approach has been particularly useful, for example, for uterine cervical cancer. The great difficulty of applying this approach to melanoma screening is the documentation of the extent of screening in different communities. An approach, however, could be made by systematic surveys, and if large communities could be assessed that differ considerably in screening, useful information might be gained.

A third option in evaluation is the use of case–control methods — comparing subjects who die from melanoma or subjects with deeply invasive melanoma with controls drawn from the eligible population. In addition to the difficulties of design inherent in the case–control approach when applied to cancer screening, the particular difficulty again is the adequate and unbiased estimation of the frequency and timing of previous skin examinations. If this information has to be obtained from interview with the subjects or with surviving relatives, the problems of error and bias may be considerable. There might be better possibilities to mount a case–control study in a population that might have more accurate documentation of past skin examinations, perhaps in the context of a health medicine organization or some other situation with particularly good medical records. However, the potential value of the case–control approach still holds. The case–control approach also has the potential of demonstrating whether screening results in the overdiagnosis of thin, nonprogressive lesions, and proposals for such case–control studies have been produced (Elwood, unpublished). Berwick et al. [73] have compared reported screening in 123 incident cases of "lethal" melanoma in Connecticut, and in control subjects obtained by random digit dialing. Preliminary results show a significant protective effect of skin awareness, and a nonsignificant protective effect for skin self-examination, adjusted for sex, age, sun exposure, total nevi,

skin color, education, and family history of skin cancer. The response rate in those with deeply invasive melanoma was less than in other subjects, and there was a reduced odds ratio between all melanoma and the controls, which suggests a self-selection effect for screening among all melanoma patients, or bias in response to the questions, or selection effects in the control group.

11. New methods of skin screening

The method of skin screening discussed so far is clinical examination of the skin, without any technological aids. Epiluminescence microscopy (ELM) is a method of visualization of the skin using a hand-held magnifier placed on the skin after application of a thin layer of mineral oil. ELM provides visibility though the upper epidermis and a modest degree of magnification. The instrument can be supplemented by a camera or electronic imaging system. Kenet et al. [74] have reviewed the potential of ELM in improving the diagnosis of suspected pigmented lesions, describing patterns that appear most characteristic of melanoma (such as a multicomponent pattern, nodular pattern, pseudopods, radial-streaming and blue-grey areas) as well as some patterns more characteristic of benign pigmented lesions (such as a saccular pattern, globular pattern, or comedolike openings). Steiner et al. [75] have shown that the diagnostic accuracy for pigmented Spitz nevi (which can be confused with melanoma) can be improved by the use of ELM. However, the use of ELM even as a diagnostic aid for lesions that have already been recognized as suspicious is in its early stages and is a long way from being useful as a screening device.

Another method is the use of computerized image analysis methods to distinguish between pigmented lesions that are likely to be malignant and those that are likely to be benign. Preliminary work on the use of such systems is being carried out in several countries, using computerized image-analysis techniques applied either to standardized photographs or directly to patients' lesions using a camera. Green et al. [76] have developed a system using a color video camera and a computer program for image analysis that assesses the size, color, shape, and boundary of a skin lesion and uses discriminant analysis of past results to provide a binary diagnostic decision between melanoma and nonmelanoma [76]. In a study of 164 lesions excised from 129 unselected patients, there were 18 melanomas. Sixteen of these were correctly classified by the computer system, as compared to 15 correctly diagnosed clinically; the computer system gave 16 false-positive diagnoses, compared to 22 clinical false-positive diagnoses; thus the system has been shown to be as or more sensitive and more specific than clinical assessment. Both the computer system and the clinical assessment misdiagnosed the same two melanomas, which were small level-one lesions. Several systems have been developed, and some are now commercially available from major manufacturers. The development potential of such techniques seems very considerable, since more sophisticated

artificial intelligence could be used, using neural networks and learning capabilities rather than statistical analysis alone, and the system could include historical data such as the length of time the lesion has been present and history of changes as well as current information. The great potential of such systems is that if they can approximate the clinical skills of an experienced dermatologist, they could make that level of skill available in primary care. This might particularly help in dealing with the problem of nonspecificity of current clinical guidelines and the very high frequencies of removal of innocent lesions in areas with a high incidence and public awareness of melanoma.

12. Conclusions

Systematic screening of the general population for malignant melanoma is not recommended at present because there is no adequate evidence available of its effectiveness. However, there is evidence of the *potential* effectiveness of such screening. The evaluation of screening methods for malignant melanoma in the general population is therefore a high priority for research, and the general recommendation will need to be reviewed in the light of new research results. This recommendation applies both to screening by skin examinations by a doctor, and also to self-screening.

Screening of individuals at considerable increased risk of melanoma (because of factors such as the combination of the family history of melanoma and a personal history of dysplastic nevi) is recommended, since evidence suggests that such screening is likely to produce an improved outcome. This recommendation, however, applies to small groups of particularly high-risk subjects who should receive careful surveillance by a specialist or specialist unit. The temptation to include in this group subjects who may have less well-defined risk factors from melanoma, such as dysplastic or multiple nevi without family history, should be discouraged.

The above recommendations refer to screening, that is, conducting systematic skin examinations on selected subjects. In contrast, there is reasonable evidence that public and professional education programs aimed at increasing public awareness of the early signs of melanoma, followed by good primary care assessment and referral service, are effective in reducing melanoma morbidity and mortality. Such programs can therefore be supported, although a more systematic approach with appropriate evaluation is advisable.

The evaluation of melanoma screening is not easy, because of the possibilities of selection bias, lead time effects, and the diagnosis of clinically and pathological abnormal lesions with a benign natural history. Only a randomized trial will adequately deal with these issues. In view of the importance of melanoma and the consequences of undertaking screening on a large scale if doing so is in fact of little benefit, randomized trials should be given high priority. Other methods of evaluation will also be useful, but will have substantial limitations in their ability to assess ultimate benefit.

Acknowledgments

This manuscript was prepared with the skilled assistance of Ms. S. McCord and Mrs. J. Jopson. The work was supported by the Cancer Research Trust.

References

1. Elwood JM. 1991. Screening and early diagnosis for melanoma in Australia and New Zealand. In Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.), Cancer Screening. Cambridge: Cambridge University Press, pp. 243–256.
2. Ellman R. 1991. Screening for melanoma in the U.K. In Miller AB (ed.), Cancer Screening. Geneva: UICC, pp. 257–266.
3. Koh HK, Geller AC, Miller DR, Lew RA. 1991. Screening for melanoma/skin cancer in the United States. In Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC (eds.), Cancer Screening. Cambridge: International Union Against Cancer, Cambridge University Press, pp. 267–278.
4. Elwood JM. 1994. Screening for melanoma and options for its evaluation. J Med Screening 1:22–38.
5. Elder DE. 1995. Skin cancer. Melanoma and other specific nonmelanoma skin cancers. Cancer 75:245–256.
6. Roush GC, McKay L, Holford TR. 1992. A reversal of the long-term increase in deaths attributable to malignant melanoma. Cancer 69:1714–1720.
7. Scotto J, Pitcher H, Lee JAH. 1991. Indications of future decreasing trends in skin-melanoma mortality among whites in the United States. Int J Cancer 49:490–497.
8. Chen Y, Zheng T, Holford TR, Berwick M, Dubrow R. 1994. Malignant melanoma incidence in Connecticut (United States): time trends and age-period-cohort modeling by anatomic site. Cancer Causes Control 5:341–350.
9. Elwood JM. 1992. Primary prevention and early diagnosis of melanoma in Canada: some options for development. Chron Dis Can 13:102–110.
10. MacLennan R, Green AC, McLeod GRC, Martin NG. 1992. Increasing incidence of cutaneous melanoma in Queensland, Australia. J Natl Cancer Inst 84:1427–1432.
11. Jones ME, Shugg D, Dwyer T, Young B, Bonett A. 1992. Interstate differences in incidence and mortality from melanoma: a re-examination of the latitudinal gradient. Med J Aust 157:373–378.
12. Thörn M, Sparén P, Bergström R, Adami H. 1992. Trends in mortality rates from malignant melanoma in Sweden 1953–1987 and forecasts up to 2007. Br J Cancer 66:563–567.
13. Thorn M, Ponten F, Bergstrom R, Sparen P, Adami HO. 1994. Trends in tumour characteristics and survival of malignant melanoma 1960–84: a population-based study in Sweden. Br J Cancer 70:743–748.
14. Urist MM, Karnell LH. 1994. The National Cancer Data Base. Report on melanoma. Cancer 74:782–788.
15. Morris J. 1994. Screening for malignant melanoma. J Med Screening 1:2.
16. MacKie RM, Hole D. 1992. Audit of public education campaign to encourage earlier detection of malignant melanoma. Br Med J 304:1012–1015.
17. Elwood JM. 1993. Recent developments in melanoma epidemiology, 1993. Melanoma Res 3:149–156.
18. Burton RC, Armstrong BK. 1994. Recent incidence trends imply a non-metastasizing form of invasive melanoma. Melanoma Res 4:107–113.
19. Burton RC, Coates MS, Hersey P, et al. 1995. An analysis of a melanoma epidemic. Int J Cancer 55:765–770.
20. Brown L, Palmer PH. 1991. Melanoma incidence in Tauranga 1980–9. N Z Med J 908:109–111.

21. Elwood JM, Glasgow H. 1993. The Prevention and Early Detection of Melanoma in New Zealand. Wellington: Cancer Society of New Zealand and Department of Health, pp. 1–44.
22. Elwood JM. 1990. Screening programmes in disease control. In McNeil JJ, King RWF, Jennings GL, Powles JW (eds.), A Textbook of Preventive Medicine. Melbourne: Edward Arnold, pp. 23–44.
23. American Cancer Society. 1991. Guidelines for the cancer-related checkup: recommendations and rationale. Atlanta: American Cancer Society.
24. American Academy of Dermatology Committee on Guidelines of Care. 1992. Guidelines of care for nevi 1 (nevocellular nevi and seborrheic keratoses). J Am Acad Dermatol 26:629–631.
25. Friedman RJ, Rigel DS, Kopf AW. 1985. Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J Clin 35:130–151.
26. Canadian Task Force on the Periodic Health Examination. 1984. The periodic health examination: 2. 1984 update. Can Med Assoc J 130:1278–1285.
27. U.S.Preventive Services Task Force. 1989. Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Baltimore: Williams & Wilkins.
28. Marks R. 1991. Screening for melanoma (letter). Med J Aust 154:707.
29. International Union Against Cancer (UICC). 1992. Melanoma Control Manual. Geneva: UICC.
30. Kramer BS. 1995. NCI state of the art statements on cancer screening: screening for skin cancer. In Greenwald P, Kramer BS, Weed DL (eds.), Cancer Prevention and Control. New York: Marcel Dekker, pp. 746–747.
31. National Institutes of Health. 1992. Diagnosis and treatment of early melanoma. NIH Consensus Development Conference: Consensus Statement.
32. Girgis A, Campbell EM, Redman S, Sanson-Fisher RW. 1991. Screening for melanoma: a community survey of prevalence and predictors. Med J Aust 154:338–343.
33. Queensland Health. 1991. Enhancing the early detection of melanoma in Queensland. Information circular 23:1–7.
34. Koh HK. 1991. Cutaneous melanoma. N Engl J Med 325:171–182.
35. Jelfs PL, Giles G, Shugg D, et al. 1994. Cutaneous malignant melanoma in Australia, 1989. Med J Aust 161:182–187.
36. Koh HK, Caruso A, Gage I, et al. 1990. Evaluation of melanoma/skin cancer screening in Massachusetts. Cancer 65:375–379.
37. Rampen FHJ, van Huystee BEWL, Kiemeney LALM. 1991. Melanoma/skin cancer screening clinics: experiences in the Netherlands. J Am Acad Dermatol 25:776–777.
38. Brandberg Y, Bolund C, Micheleson H, et al. 1993. Psychological reactions in public melanoma screening. Eur J Cancer 29:860–863.
39. Brandberg Y, Bergenmar M, Bolund C, Månsson-Brahme E, Ringborg U, Sjxdén P. 1992. Psychological effects of participation in a prevention programme for individuals with increased risk for malignant melanoma. Eur J Cancer 28A:1334–1338.
40. Koh HK, Lew RA, Prout MN. 1989. Screening for melanoma/skin cancer; theoretic and practical considerations. J Am Acad Dermatol 20:159–172.
41. Kopf AW, Mintzis M, Bart RS. 1975. Diagnostic accuracy in malignant melanoma. Arch Dermatol 111:1291–1292.
42. Rigel DS, Friedman RJ, Kopf AW, et al. 1986. Importance of complete cutaneous examination for the detection of malignant melanoma. J Am Acad Dermatol 14:857–860.
43. Leffell DJ, Berwick M, Bolognia J. 1993. The effect of pre-education on patient compliance with full-body examination in a public skin cancer screening. J Dermatol Surg Oncol 19:660–663.
44. Fitzpatrick TB, Rhodes AR, Sober AJ, Mihm CM Jr. 1988. Primary malignant melanoma of the skin: the call for action to identify persons at risk; to discover precursor lesions; to detect early melanomas. In Elwood JM (ed.), Naevi and Melanoma: Incidence, Interrelationships and Implications; Pigment Cell no. 9. Karger: Basel, pp. 110–117.
45. MacKie RM. 1990. Clinical recognition of early invasive malignant melanoma: looking for changes in size, shape, and colour is successful. Br Med J 301:1005–1006.

46. Rhodes AR. 1995. Public education and cancer of the skin: what do people need to know about melanoma and nonmelanoma skin cancer? Cancer 75:613–636.
47. Leffell DJ, Chen Y, Berwick M, Bolognia JL. 1993. Interobserver agreement in a community skin cancer screening setting. J Am Acad Dermatol 28:1003–1005.
48. DeCoste SD, Stern RS. 1993. Diagnosis and treatment of nevomelanocytic lesions of the skin. Arch Dermatol 129:57–62.
49. Gruber SB, Roush GC, Barnhill RL. 1993. Sensitivity and specificity of self-examination for cutaneous malignant melanoma risk factors. Am J Prev Med 9:50–54.
50. Lawson DD, Moore DH, Schneider JS, Sagebiel RW. 1994. Nevus counting as a risk factor for melanoma: comparison of self-count with count by physician. J Am Acad Dermatol 31:438–444.
51. Byles JE, Hennrikus D, Sanson-Fisher RW, Hersey P. 1994. Reliability of naevus counts in identifying individuals at high risk of malignant melanoma. Br J Dermatol 130:51–56.
52. Hennrikus D, Girgis A, Redman S, Sanson-Fisher RW. 1991. A community study of delay in presenting with signs of melanoma to medical practitioners. Arch Dermatol 127:356–361.
53. MacKie RM. 1993. What are education and screening and diagnosis and prevention of melanoma achieving? Melanoma Research 3:7 (abstract).
54. Bulliard J, Raymond L, Levi F, et al. 1992. Prevention of cutaneous melanoma: an epidemiological evaluation of the Swiss campaign. Rev Epidemiol Sante Publique 40:431–438.
55. Koh HK, Geller AC, Miller DR, Caruso A, Gage I, Lew RA. 1991. Who is being screened for melanoma/skin cancer? Characteristics of persons screened in Massachusetts. J Am Acad Dermatol 24:271–277.
56. McGee R, Elwood M, Williams S, Lowry F. 1994. Who comes to skin checks? N Z Med J 107:58–60.
57. Rampen FHJ, Berretty PJM, van Huystee BEWL, Kiemeney LALM, Nijs CHHM. 1993. Lack of selective attendance of participants at skin cancer/melanoma screening clinics. J Am Acad Dermatol 29:423–427.
58. Kricker A, English DR, Randell PL, et al. 1990. Skin cancer in Geraldton, Western Australia: a survey of incidence and prevalence. Med J Aust 152:399–407.
59. Reynolds P, Austin DF. 1985. Epidemiologic based screening strategy for malignant melanoma. Prog Clin Biol Res 156:245–254.
60. Moore D, Martin K, Schneider J. 1993. Aggressive awareness campaign can reduce melanoma mortality. Melanoma Res 3:55 (abstract).
61. Kraemer KH, Tucker M, Tarone R, Elder DE, Clark WH Jr. 1986. Risk of cutaneous melanoma in dysplastic nevus syndrome types A and B (letter). N Engl J Med 315:1615–1616.
62. Greene MH, Clark WH, Tucker MA, et al. 1984. Managing the dysplastic naevus syndrome. Lancet i:166–167.
63. MacKie RM, McHenry P, Hole D. 1993. Accelerated detection with prospective surveillance for cutaneous malignant melanoma in high risk groups. Lancet 341:1618–1620.
64. Vasen HFA, Berman W, Haeringen A, et al. 1989. The familial dysplastic nevus syndrome. Natural history and the impact of screening on prognosis — a study of nine families in the Netherlands. Eur J Cancer Clin Oncol 25:337–341.
65. Elwood JM, Cooke KR, Coombs BD, Cox B, Hand JE, Skegg DCG. 1988. A strategy for the control of malignant melanoma in New Zealand. N Z Med J 101:602–604.
66. Duray PH, DerSimonian R, Barnhill R, et al. 1992. An analysis of interobserver recognition of the histopathologic features of dysplastic nevi from a mixed group of nevomelanocytic lesions. J Am Acad Dermatol 27:741–749.
67. Smoller B, Egberg B. 1992. Dysplastic nevi can be diagnosed and graded reproducibility: a longitudinal study. J Am Acad Dermatol 27:741–749.
68. English DR, Armstrong BK. 1988. Identifying people at high risk of cutaneous malignant melanoma: results from a case–control study in Western Australia. Br Med J 296:1285–1288.
69. Marrett LD, King WD, Walter SD, From L. 1992. Use of host factors to identify people at high risk for cutaneous malignant melanoma. Can Med Assoc J 147:445–453.

70. MacKie RM, Freudenberger T, Aitchison TC. 1989. Personal risk-factor chart for cutaneous melanoma. Lancet 2:487–490.
71. Goldenhersh MA. 1993. Melanoma screening: critique and proposal. J Am Acad Dermatol 28:642–644.
72. Koh HK, Geller AC, Miller DR, Lew RA. 1993. Early detection of melanoma: an ounce of prevention may be a ton of work. J Am Acad Dermatol 28:645–647.
73. Berwick M, Dubin N, Roush G, Barnhill R. 1993. Early detection and lethal melanoma in Connecticut: a preliminary analysis. In Gallagher RP, Elwood JM (eds.), Epidemiological Aspects of Cutaneous Malignant Melanoma. Boston: Kluwer, pp. 265–271.
74. Kenet RO, Kang S, Kenet BJ, et al. 1993. Clinical diagnosis of pigmented lesions using digital epiluminescence microscopy. Arch Dermatol 129:157–175.
75. Steiner A, Pehamberger H, Binder M, et al. 1992. Pigmented Spitz nevi: improvement of the diagnostic accuracy by epiluminescence microscopy. J Am Acad Dermatol 27:697–701.
76. Green A, Martin N, Pfitzner J, O'Rourke M, Knight N. 1994. Computer image analysis in the diagnosis of melanoma. J Am Acad Dermatol 31:958–964.

11. Screening for Neuroblastoma

Mark L. Bernstein and William G. Woods

1. Introduction

Screening for a malignant disease is an attempt to find an asymptomatic and curable disease before it develops into one that is both symptomatic and incurable. The success of such a venture depends upon the sensitivity and specificity of the screening tool used, the extent to which the target population participates in the screening program, the ease with which the asymptomatic disease may be cured, and the likelihood that the asymptomatic disease will progress into an advanced-stage malignancy. Neuroblastoma is likely to be the only pediatric malignancy subject to a screening trial, since the tumor commonly secretes chemicals that can be detected in urine, the disease is of relatively high frequency (exceeded only by hypothyroidism among diseases currently screened for in the early newborn period), and, when present in advanced form in older infants, it is a disease that is extremely difficult to cure.

2. Neuroblastoma: a brief overview of an enigmatic disease

Neuroblastoma is the most common extracranial solid tumor of childhood, affecting approximately 1:7200 children under the age of five years [1]. It has its origin within cells of the neural crest, and so can arise anywhere in the sympathetic nervous system, from the adrenal gland to the paravertebral sympathetic chain to the organ of Zuckerkandel. At least 90% of affected children are diagnosed before the age of five years, almost all before the age of eight years. Thus, the tumor is likely of congenital origin. Young age at diagnosis (<24, and particularly <12 months) and an early stage at the time of diagnosis have been the longest-known, most reliable favorable prognostic features. Various staging systems have been employed to describe the extent of disease at the time of presentation. A comparative description of these staging systems is reproduced in table 1.

In young infants, even extensive disease can undergo spontaneous regression. This is IV-S (S for special) disease, with a small primary tumor, and metastatic disease involving any combination of skin, bone marrow, and liver,

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

Table 1. Comparison of staging system for neuroblastoma

CCSG system	POG system	International system
Stage I. Tumor confined to the organ or structure of origin.	Stage A. Complete gross resection of the primary tumor, with or without microscopic residual disease. Intracavitary lymph nodes not adhered to the primary tumor must be histologically free of tumor. Nodes adhered to the surface of or within the primary may be positive.	Stage 1. Localized tumor confined to the area of origin; complete gross excision, with or without microscopic residual disease; identifiable ipsilateral and contralateral lymph nodes negative microscopically.
Stage II. Tumor extending in continuity beyond the organ or structure of origin, but not crossing the midline. Regional lymph nodes on the ipsilateral side may be involved.	Stage B. Grossly unresected primary tumor. Nodes and nodules the same as in stage A.	Stage 2A. Unilateral tumor with incomplete gross excision; identifiable ipsilateral and contralateral lymph nodes negative microscopically.
		Stage 2B. Unilateral tumor with complete or incomplete gross excision; with positive ipsilateral regional lymph nodes; identifiable contralateral lymph nodes negative microscopically.
Stage III: Tumor extending in continuity beyond the midline. Regional lymph nodes may be involved bilaterally.	Stage C. Complete or incomplete resection of primary. Intracavitary nodes not adhered to primary must be histologically positive for tumor. Liver as in stage A.	Stage 3. Tumor infiltrating across the midline with or without regional lymph node involvement; or, unilateral tumor with contralateral regional lymph node involvement; or, midline tumor with bilateral lymph node involvement.
Stage IV. Remote disease involving the skeleton, bone marrow, soft tissue, and distant lymph node groups (see stage IV-S).	Stage D. Dissemination of disease beyond intracavitary nodes (i.e., extracavitary nodes, liver, skin, bone marrow, bone, etc).	Stage 4. Dissemination of tumor to distant lymph nodes, bone, bone marrow, liver, and/or other organs (except as defined in stage 4S).
Stage IV-S. As defined in stage I or II, except for the presence of remote disease confined to the liver, skin, or marrow (without bone metastases).	Stage DS. Infants <1 yr of age with stage IV-S disease (see CCSG).	Stage 4S. Localized primary tumor as defined for stage 1 or 2 with dissemination limited to liver, skin, and/or bone marrow.

From Brodeur G, [6], with permission.

but specifically not skeletal bone. The existence of this unusual form of disease suggests that less widespread disease may also undergo spontaneous maturation and regression. In fact, a variety of studies over time have suggested that treatment beyond partial surgical removal may not be necessary for early-stage disease in young infants. Moreover, small foci of neuroblasts may be seen in the adrenal glands of fetuses and young infants who die of other causes, with a frequency as high as 1:250 [2,3]. These "neuroblastomas-in-situ" presumably mature normally, without causing disease in the host. Thus, if neuroblastomas-in-situ produce catecholamine metabolites in sufficient quantity to be detected by a screening test, a potentially confounding variable may be introduced into a screening program, with a large number of children detected with a condition that requires no therapy.

More recently, the biology of neuroblastoma has also been found to be a very important determinant of prognosis. Two of the most important features are the DNA content of the malignant neuroblast and the presence or absence of amplification of the N-myc oncogene. If the DNA content of a normal cell containing 46 chromosomes is arbitrarily defined as 1.0, then the most favorable neuroblastomas are near-triploid, with a DNA content of 1.25–1.75. Diploidy is unfavorable, as is tetraploidy, or other unusual ploidy. These ploidy findings are of particular prognostic significance in children less than 12 months of age at the time of diagnosis. Absence of amplification of the N-myc oncogene is clearly prognostically favorable when compared with its presence.

These prognostic features interact; that is, children over 12 months of age and those with advanced disease are much more likely to have either an unfavorable DNA index, an amplified N-myc oncogene, or both. Moreover, most of those with an amplified N-myc oncogene also have an unfavorable DNA index. Tumors that present clinically have consistent biologic features both at different disease sites and over time. Those that have unamplified N-myc at diagnosis also have unamplified N-myc at second-look surgery or at relapse. DNA content also remains constant. Therefore, there are no clear examples of biologic progression of disease for clinically diagnosed cases, although it is unclear if such progression occurs in the preclinical period [4,5].

Approximately one third of children with neuroblastoma present with disseminated disease at more than 12 months of age [1]. Therapy of such children has been particularly disappointing, with little improvement in a 10% to 15% survival rate over the past 25 years. This is true even for children who undergo bone marrow transplantation. The intensive preparative regimens seem able to prolong remission duration without necessarily improving the ultimate cure rate. The existence of this substantial subpopulation of hard-to-treat children within the group of children with neuroblastoma has been a driving force behind the institution of screening programs [6–8]. In addition to the difficulty of treating children with advanced-stage neuroblastoma, the other feature rendering a screening program feasible is the presence in the urine of two of the catecholamine metabolites in 85% to 90% of children at the time of diagnosis. These two commonly sought metabolites are homovanillic acid

(HVA), a product of dopamine metabolism, and vanillylmandelic acid (VMA), a product of norepinephrine and epinephrine metabolism. Since dopamine beta hydroxylase is required to convert dopamine to norepinephrine, the presence of VMA is consequently a marker of a more mature tumor. Both HVA and VMA are relatively stable, and can be eluted from a urine-soaked air-dried filter paper. Moreover, when normalized to creatinine, there is little diurnal variation in their excretion, so that a random urine sample is sufficient for analysis [9–12].

3. The early Japanese experience

In 1973, Sawada et al. instituted a pilot screening program in Kyoto, analyzing air-dried urine-soaked filter papers by elution and qualitative spot analysis using the LaBrosse test [13]. Children whose urine was positive or intermediate were asked to submit two additional samples, and, if those were still abnormal, to undergo physical examination, quantitative analysis of HVA and VMA, and radiographic examination of the chest and abdomen. Over a six-year period, 78,831 children were screened (an estimated 62.5% of eligible infants in the region). Fifty-two had a positive screen (1 in 1515). Four of these children were found to have neuroblastomas, all of which were successfully treated. Only one child who had a negative screen subsequently developed a neuroblastoma, and was still VMA-spot negative at the time of presentation with cerebellar ataxia.

Subsequent extension of this series to other communities in Japan showed similar results [14]. Using the same qualitative LaBrosse spot test, 281,939 infants were screened from 1981–1983 (with the exception of Kyoto, as discussed above). Two hundred and sixty-four children were referred for examination (1 in 1068), of whom 16 (1 in 16) were found to have tumors. Fifteen were alive at the time of the report. Six children who were screen-negative were found to have neuroblastomas between the ages of 14 and 29 months. Their catecholamine status at the time of diagnosis, as well as their outcome, was not stated in the report.

A technical improvement in the screening technique was introduced in Sapporo City, where quantitative analysis of both HVA and VMA by high-performance liquid chromatography (HPLC) was substituted for qualitative spot analysis [15]. An approximate 75% of the population of 97,862 infants was screened. Fifteen infants with neuroblastoma were detected (false-positive rate not given). Fourteen of the 15 were in remission at the time of the report, with only one death. Four additional cases, all of whom were catecholamine-positive at the time of diagnosis, were negative at the time of screening. Overall, the effect appeared dramatic, with an increase in the survival rate in Sapporo City from 21% to 87%, while the survival rate in the surrounding Hokkaido Prefecture, the remainder of the island, remained unchanged over the same time period: 21% in the years 1969–1980, 28% from 1981–1984.

4. Questions concerning the early Japanese experience

A reanalysis of the early Japanese data [16] suggested that significantly more children under one year of age had been diagnosed as having neuroblastoma in Sapporo City following the institution of screening, while there was a nonsignificant change in the number of diagnoses in older children. Subsequent data provided by one of the investigators confirmed this impression of increased incidence [17]. The surrounding Prefecture of Hokkaido showed no change over the same time interval. Moreover, most of the early infantile cases diagnosed in Sapporo City were of favorable biology and small size at the time of diagnosis (for further detail, see below), further supporting the possibility of overdiagnosis. Overall mortality statistics were not reliably available, although the numbers supplied suggested an overall decrease in mortality from neuroblastoma in Japan even before the institution of nationwide mandatory screening at six months of age.

5. The Quebec study

5.1. Feasibility

The intriguing preliminary results from Japan, and the existence in the province of Quebec of an early metabolic screen with a high participation rate and an attendant infrastructure [18], made the idea of a controlled trial of screening in the province of Quebec, using other North American populations as control groups for population-based incidence and mortality, an attractive one. We first determined that the population would be receptive to the idea of a "tumor screen" in early childhood, and that, with an annual birth rate of 80,000 to 100,000 in the province and an incidence of at least 1:10,000, sufficient cases could be detected so as to make such a trial feasible [19]. We then retrospectively analyzed the population-based incidence and mortality of neuroblastoma in Quebec and in one of the proposed control groups, the Greater Delaware Valley, a 31-county region including Philadelphia with a long-established population and pathology-based tumor registry [1]. The incidence and mortality were not, in fact, statistically different in the two regions. In Quebec, the annual incidence rates for neuroblastoma were 11.32 per million children under 15 years of age and 27.47 per million children under five years of age (cumulative incidence approximately 1:7300 for children under five years of age), whereas in the Greater Delaware Valley the corresponding numbers were 10.58 per million for ages 0–14 and 28.03 per million for children aged 0–4 years (cumulative incidence was approximately 1:7150 for the younger children). The annual mortality rates in the two regions were 9.04 per million children in Quebec aged 0–4 years, 4.82 per million for children 0–14 years (40 total deaths in the time period from 1981–1986, with 25 in the younger group) as compared with 9.15 per million younger children in the

Greater Delaware Valley and 4.96 per million for the entire group (52 deaths overall, 32 in the children aged 0–4 years). Interestingly, the overall survival for all children with neuroblastoma in both regions was approximately 55%, substantially better than the 28% that had been reported from Hokkaido Prefecture for roughly the same time period. This led to our calculation of the need for a five-year study, accruing approximately 450,000 births with a minimum participation rate in the screening program of 75%, to afford us an 80% probability of detecting an improvement in the survival rate of 20% (from 55% to 75%) as a surrogate endpoint, while at the same time we monitored overall population-based incidence and mortality. Four control groups were chosen to provide comparison data for population-based incidence and mortality: the Greater Delaware Valley, the Province of Ontario, Minnesota, and Florida. Similar calculations, performed independently, came to a similar conclusion concerning the size of a study required to demonstrate a health benefit of screening [17].

6. Implementation of the Quebec study

6.1. Design

Air-dried urine-soaked filter papers were mailed to the urine screening laboratory in Sherbrooke at the parents' expense at two time points — three weeks and six months of age. Three weeks was chosen since it is the time at which the already established urine metabolic screen is performed [18], six months because of the extensive preliminary Japanese experience at that age, as well as the tie-in to routine immunization, facilitating a reminder to the family. The Quebec Government initially agreed to provide a reminder mailed along with the six-month family allowance check. As electronic deposit became increasingly common, this strategy was no longer possible. Instead, publicity posters in physicians offices and other health care facilities, reminders in the media, and, finally, a mailing to families who had not submitted a six-month sample were employed.

Screening was performed in two stages. First, a semiquantitative screen was performed in Sherbrooke using thin-layer chromatography, a technique with which the Sherbrooke group has extensive experience [20]. Borderline or abnormal samples were then sent to Minneapolis for quantitation using capillary gas chromatography and mass spectroscopy [21]. Children whose urinary excretion of HVA or VMA was abnormally elevated were asked to submit a second air-dried urine-soaked filter paper. Those with two abnormally elevated results were then asked to be brought to one of the four academic health science centers in the Province of Quebec — Hôpital Ste.-Justine (Université de Montréal), the Montreal Children's Hospital (McGill University), le Centre Hospitalier de l'Université Laval (Quebec City), or le Centre Hospitalier de l'Université Sherbrooke. There, the children underwent a

physical examination, a repeat quantitative test for urinary HVA and VMA, and a radiologic examination consisting of a chest radiograph and an abdominal ultrasound. A computed tomogram of the abdomen was also performed on children with two positive six-month screens. Children found to have a mass were then staged and treated according to contemporary Pediatric Oncology Group protocols. Biologic studies of tumor and normal tissue were also undertaken. Data from each of the four population-based tumor registries were reviewed as they became available.

6.2. Results

The eligible population cohort consisted of all children born in the Province of Quebec from May 1, 1989, through April 30, 1994. This population numbered 470,236. Ninety-one percent participated in the screen at three weeks of age, for a total of 430,000 samples. As of August, 1994 (63 months of observation, and four months before the end of screening for neuroblastoma in the province), approximately 75% of the eligible six-month-olds had submitted a urine sample, representing 325,000 children. The combination of thin-layer chromatography and gas chromatography/mass spectroscopy was remarkably specific. Seventy-nine children were referred for evaluation, 38 after two positive three-week screens and 41 after two positive six-month screening tests. Forty were diagnosed as having neuroblastoma, 18 after the three-week screen and 22 after the six-month screen. Moreover, three were felt on imaging studies to possibly have regressing neuroblastomas. Even disregarding these cases, this gives the screen a positive predictive value of 51%, and a specificity of more than 99.99% at both ages. Differently stated, approximately one child in 11,000 was referred for evaluation at three weeks of age, whereas one in 8000 was referred at six months of age. Half of those evaluated had neuroblastomas. Of note, all of those with neuroblastomas had elevated urinary VMAs, whereas only approximately three quarters also had elevated HVAs.

The most dramatic effect of the screening program was on the incidence of neuroblastoma in young infants (table 2). Seventy-five cases of neuroblastoma were observed in children less than 12 months of age, compared with the 29 expected on the basis of the historical experience in the Greater Delaware Valley and data from the Surveillance, Epidemiology, and End Results Program of the National Cancer Institute (U.S.A.) [22]. The standardized incidence ratios for the two age groups, 0–6 months and 6–12 months, are respectively 2.84 (95% confidence interval, 2.07–3.73) and 2.33 (95% CI, 1.53–3.29). It is extremely unlikely, therefore, that the increased incidence was a chance finding. Early detection of tumors might, in fact, be expected to raise the incidence of disease in young infants, while correspondingly decreasing it in children over 12 months of age, especially those in the one- to four-year-old age group. Unfortunately, no such decrease was observed, with the number of observed cases falling well within the 95% likelihood of those expected (25 observed, 18 expected). At the same time, the number of cases observed in the

Table 2. Standardized incidence ratio (SIR) in Quebec (63 months observation)

Age at diagnosis (months)	Person-years to date	Annual incidence per million	Observed/ expected	SIR	95% CI
0–5	227,905	72.6	47/16.6	2.84[a]	2.07–3.73
6–11	206,829	58.1	28/12.0	2.33[a]	1.53–3.29
12–17	183,662	30.8	8/5.7	1.41	0.59–2.59
18–23	161,027	29.1	9/4.7	1.92	0.85–3.41
24–29	136,889	13.8	4/1.9	2.12	0.53–4.76
30–35	113,234	24.2	2/2.7	0.73	0.06–2.13
36–41	88,685	17.5	1/1.6	0.64	0.00–2.58
42–47	65,053	14.0	1/0.9	1.10	0.00–4.39
48–53	43,240	22.9	1/1.0	1.01	0.00–4.04
54–59	21,636	10.6	1/0.2	4.36	0.00–17.44
Summary					
0–59	1,248,160		102/47	2.16[a]	1.75–2.61

[a] $p < 0.05$.

Table 3. International staging of patients detected in Quebec (63 months observation)

Stage	Preclinically detected tumors ($N = 40$)	Clinically detected tumors ($N = 62$)
1	9	18
2A	1	3
2B	10	4
4S	7	9
3	5	5
4	8	23

two control groups with the most rapid ascertainment, Minnesota and Ontario, was exactly the number expected — 99 cases observed versus 100 cases expected, from a birth cohort of approximately 200,000 per year (65,000 in Minnesota, 135,000 in Ontario).

All stages of disease are included among the 40 children diagnosed as having neuroblastoma through screening (table 3). Moreover, for the entire group of 102 children, the more advanced stages of disease were fully represented — 40 observed cases in stages III and IV for the group aged 0–4 years versus 27 expected, 18 of them between the ages of 1 and 4 years (14 expected).

Biologically, however, almost all the tumors with unfavorable ploidy, and all those with an amplified N-myc oncogene, fell in the clinically detected group (table 4). Similarly, almost all the tumors histologically classified as "unfavorable" in the Shimada classification [6,7] were in the clinically diagnosed group. Only one tumor of unfavorable histology was found in the preclinically detected group. It was a stage I lesion at the time of diagnosis [23].

Table 4. Biologic results to date in Quebec (63 months observation)

	Preclinical detection	Clinical detection	Total
1. Shimada Histologic Classification			
Favorable	40	41	81
Unfavorable	1	14	15
2. DNA Ploidy			
Diploid (1.00)	0	11	11
Hyperdiploid (1.09)	0	2	2
Triploid (1.24–1.72)	28	29	57
Tetraploid (1.98–2.16)	0	3	3
Pentaploid (2.29–2.53)	1	2	3
Mixed aneuploid	6	8	14
Mixed diploid/aneuploid	3	5	8
Hypodiploid	1	0	1
3. N-*myc* Gene Amplification			
Single Copy	37	46	83
Amplified	0	9	9

Table 5. Neuroblastoma deaths in Quebec (59 months observation)

ID#	Age (mos)	INSS stage	Screened	Secretion status	Biology (N-myc; ploidy)	Histology
L17	1	4S	(Yes)[a]	(+)	Good	Unfavorable
J35	2	4S	(Yes)[a]	(+)	Good	N.A.
M17	9	4	No	(+)	Poor	N.A.
M19	12	4	Yes	(−)	Poor	N.A.
M25	14	4	Yes	(−)	Poor	Unfavorable
M27	15	4	No	(+)	Poor	Favorable
L21	22	4	Yes	(+)	Poor	Unfavorable
J39	24	4	Yes	(+)	Poor	Unfavorable
S11	24	4	Yes	(+)	Unknown	Unknown
L22	28	4	Yes	(−)	Poor	Unfavorable
J66	31	4	Yes	(+)	Good	N.A.
J84	42	4	Yes	(+)	Good	N.A.
Totals	Median 18	$N = 12$	10/12 Yes	9/12 (+)	7/11 Poor	5/6 Unfavorable

[a] Filter paper sent, not yet analyzed.

All of the 12 deaths from neuroblastoma fell within the clinically detected group (table 5). All but two of the deaths were in children with stage IV disease. Of those with stage IV disease, only two had a favorable DNA index and an unamplified N-myc oncogene. These two were the two oldest children, 31 and 42 months of age at diagnosis, in whom ploidy is of less prognostic significance. The two young infants who died of stage IVS disease had very rapidly growing abdominal masses that led to death despite intensive therapy. In the same two control groups discussed above, Minnesota and Ontario, there were 13 deaths over the same time period in a population slightly more than twice as large, but with slightly slower mortality ascertainment. It is unlikely,

however, that the number of deaths in Quebec will be less than that in the control regions.

In the entire group of 102 children, 20 were diagnosed before the age of three weeks, seven of whom had submitted filter papers that were not yet analyzed. All had elevated urinary HVA or VMA. While this may represent a "halo effect" from the publicity surrounding the urinary screening program, it may also represent a general improvement in the standard of care, since an increased number of very young infants were also diagnosed in Ontario over the same time period, when compared with historical experience.

Twelve children were catecholamine nonsecretors at the time of diagnosis (12% of the entire population; 19% of the clinically diagnosed patients). Seven children had true false-negative screening tests. That is, at the time of diagnosis of neuroblastoma, archived filter papers obtained at the time of screening were reanalyzed quantitatively by gas chromatography/mass spectroscopy and found to contain elevated levels of VMA or HVA. When compared with the upper limit of normal, these samples contained approximately 1.5 times as much VMA. No sample contained more than twice the upper limit of normal of VMA. Four children had been missed at three weeks of age. Three of those were detected on the six-month screen, and one was diagnosed clinically at four months of age. Of the three not detected on the six-month screen, the clinical diagnoses were made at the ages of 7, 8, and 53 months. Only the child diagnosed at 53 months failed to respond to initial therapy (cyclophosphamide and adriamycin) and required more intensive treatment with multiple-agent chemotherapy followed by autologous bone marrow transplantation. Despite biologically favorable disease, with a DNA index of 1.64 (that is, near triploid) and an unamplified N-myc oncogene, this child has developed progressive disease and is receiving palliative radiotherapy. The other six children are alive and well.

Other biologic parameters that will be investigated in the Quebec cohort include loss of heterozygosity at chromosome 1p, an unfavorable feature possibly associated with the loss or inactivation of a tumor suppressor gene [6,7]; TRK proto-oncogene expression, a favorable feature associated with the production of this component of the high-affinity nerve cell growth factor receptor [24]; p53 mutations in the malignant cell (a preliminary study of 38 consecutive tumors of various stages showed no significant mutational events in coding sequences, but 37 of the 38 tumors studied had no amplification of the N-myc oncogene [25]); and the predictive values of serum ferritin and neuron-specific enolase.

7. More recent Japanese experience

Bessho et al. [26] reviewed data provided to the Kanto-Ko-Shin-Etsu branch of the Japan Children's Cancer Registry. This registry is not population-based. There is no systematic pathology review. Rather, responses provided on a

questionnaire are used. When several time periods before screening was instituted were compared to time periods afterwards, the investigators concluded that the incidence of neuroblastoma seemingly increased, since prior to screening, neuroblastoma represented approximately 10% of all malignant disease reported, whereas afterwards it represented 19.3%. Moreover, eliminating all the cases diagnosed by screening gave a relative occurrence of 13.8%, similar to the figure obtained prior to screening. In addition, eliminating cases diagnosed by screening increased the proportion of children diagnosed after the age of one year from 50% to 75%. The authors concluded that the overall impact of a screening program for neuroblastoma on the success of therapy for that disease was uncertain.

Several reviews of the biology of neuroblastomas detected through the mass screening program have concluded that almost all the tumors had favorable features (unamplified N-myc oncogene, near triploidy, and favorable histology) when compared with tumors found clinically, with all deaths occuring in the clinically detected group [27–29].

The continuing occurrence of these biologically unfavorable advanced-stage tumors in older infants, which eventuate in death, has led some authors involved in the screening program to advocate screening at a later age [30], although it is uncertain whether the outcome of such unfavorable lesions could be influenced by earlier detection. The duration of the preclinical stage in such lesions is also unknown. In other words, some neuroblastomas may grow with such rapidity as to provide only a limited "window" for detection. On the other hand, the favorable biology and excellent outcome of the cases detected by screening have led to the concern that cases may be overtreated if they are treated as fully malignant, aggressive neuroblastomas [26,31]. Consequently, more limited intervention in such cases has been advocated [31]. An attempt to discriminate those requiring intensive therapy from those likely to spontaneously regress, based on the HVA/VMA ratio and the site of origin (adrenal versus extra-adrenal), has also been made [32], with, as yet, no prospective verification.

Finally, the overall mortality in Japan from neuroblastoma has been decreasing [33,34]. However, it has been decreasing in other industrialized countries as well [35–37]. In England and Finland, reported overall rates of population-based survival and mortality were very similar to the 55% survival with annual mortalities of nine per million for children aged four years or less, and five per million for children 14 or less reported for Quebec before the institution of a screening program [1]. The Japanese figure of approximately six per million for children between the ages of 1 and 4 years falls within this range [34]. Most authors feel that this improvement represents the result of improved management of children with biologically favorable disease, since, as discussed previously, children over 12 months of age at the time of diagnosis of advanced-stage disease remain very difficult to cure. If only 10% of such children are cured, and if they represent approximately one third of the population of unscreened children with neuroblastoma, then an overall

55% cure rate implies the necessity of curing 80% of all other children with neuroblastoma. This is, in fact, an achievable goal with modern treatment [6,7].

8. Other experience

Most other experience in screening for neuroblastoma has been of a pilot nature. Studies in England [38] and France [39] have demonstrated feasibility in those jurisdictions, with participation rates of 92% and, maximally, 82% during the final screening year, respectively. Several biologically favorable tumors were detected and successfully treated. In a pilot German experience [40], unusually, three unfavorable lesions were found in a total of 10 children detected from a population of 90,480 children over six months of age. One child had a stage 3 lesion of intermediate prognosis, with an N-myc copy number of 4. A second had a stage 1 tumor with a diploid karyotype and a deletion at chromosome 1p36.3. A third had a stage 1 tumor that was not only diploid with a deletion at chromosome 1p36.3, but also showed 100-fold amplification of the N-myc oncogene. This third child relapsed six months after resection of the primary tumor. This discouraging result suggests the difficulty of treating biologically unfavorable disease, even if it is discovered early. Finally, a pilot study in the United States [41] showed the likely difficulty of implementation in that country.

9. Conclusions

While the project has shown the feasibility of screening for neuroblastoma at three weeks and six months of age, it appears likely that the final result will be an increase in incidence of biologically favorable disease in young infants, with no change in the occurrence of advanced-stage, unfavorable disease or mortality in older children. Further follow-up will be necessary to prove these results. The dramatic difference in the two disease types, as determined by biologic parameters, has, however, been reinforced by the study, strengthening the conviction that less therapy is required for early-stage disease, particularly in young infants [42]. Screening at an older age, such as 18 months, is likely to be infeasible. A study demonstrating its benefit is estimated to require at least 2 million births [43]. Moreover, it is unclear, even if biologically unfavorable disease can be found earlier, whether it will be more successfully treatable. Finally, in order to be certain that the incidence of neuroblastoma has not changed over the five-year period of screening in Quebec, we hope to study the postscreening cohort of children now being born. We also hope to show the economic benefit to society of having undertaken such a project as a carefully controlled trial, rather than having instituted screening without sufficient study [44].

Acknowledgment

We thank Sandra Pisani for help with preparation of the manuscript.

References

1. Bernstein ML, Leclerc JM, Bunin G, et al. 1992. A population based study of neuroblastoma incidence, survival, and mortality in North America. J Clin Oncol 10:323–329.
2. Beckwith JB, Perrin EV. 1963. In situ neuroblastomas: a contribution to the natural history of neural crest tumours. Am J Pathol 43:1089–1104.
3. Turkel SB, Itabashi HH. 1974. The natural history of neuroblastoma cells in the fetal adrenal gland. Am J Pathol 76:225–244.
4. Brodeur GM, Hayes FA, Green AA, et al. 1987. Consistent N-myc copy number in simultaneous or consecutive neuroblastoma samples from sixty individual patients. Cancer Res 47:4248–4253.
5. Taylor SR, Blatt J, Costantino JP, et al. 1988. Flow cytometric DNA analysis of neuroblastoma and ganglioneuroma. Cancer 62:749–754.
6. Brodeur GM, Castleberry RP. 1993. Neuroblastoma. In Pizzo PA, Poplack DG (eds.), Principles and Practice of Pediatric Oncology, 2nd edition, Philadelphia: J.B. Lippincott, pp. 739–768.
7. Matthay KM. 1994. Neuroblastoma. In Pochedly C (ed.), Neoplastic Diseases of Childhood. Basel: Harwood Academic Publishers, pp. 735–778.
8. Philip T, Zucker JM, Bernard JL, et al. 1991. Improved survival at 2 and 5 years in the LMCE 1 unselected group of 72 children with stage IV neuroblastoma older than 1 year of age at diagnosis: Is cure possible in a small subgroup? J Clin Oncol 9:1045–1049.
9. Tuchman M, Morris CL, Ramnaraine ML, et al. 1985. Value of random urinary homovanillic acid and vanillylmandelic acid levels in the diagnosis and management of patients with neuroblastoma: Comparison with 24-hour urine collections. Pediatrics 75:324–328.
10. Tuchman M, Ramnaraine MLR, Woods WG, Krivit W. 1987. Three years of experience with random urinary homovanillic and vanillylmandelic acid levels in the diagnosis of neuroblastoma. Pediatrics 79:203–205.
11. Tuchman M, Robison LL, Maynard RC, et al. 1985. Assessment of the diurnal variations in urinary homovanillic and vanillylmandelic acid excretion for the diagnosis and follow-up of patients with neuroblastoma. Clin Biochem 18:176–179.
12. Tuchman M, Auray-Blais C, Ramnaraine MLR, et al. 1987. Determination of urinary homovanillic and vanillylmandelic acids from dried filter paper samples: assessment of potential methods for neuroblastoma screening. Clin Biochem 20:173–177.
13. Sawada T, Todo S, Fujita K, et al. 1982. Mass screening of neuroblastoma in infancy. Am J Dis Child 136:710–712.
14. Sawada T, Nakata T, Takasugi N, et al. 1984. Mass screening for neuroblastoma in infants in Japan: interim report of a mass screening study group. Lancet ii:271–273.
15. Nishi M, Miyake H, Takeda T, et al. 1987. Effects of the mass screening of neuroblastoma in Sapporo City. Cancer 60:433–436.
16. Goodman SN. 1991. Neuroblastoma screening data: an epidemiologic analysis. Am J Dis Child 145:1415–1422.
17. Craft AW, Parker L. 1992. Poor prognosis neuroblastoma: is screening the answer? Br J Cancer 66 (Suppl XVIII):S96–S101.
18. Lemieux B, Auray-Blais C, Giguere R, et al. 1988. Newborn urine screening experience with over one million infants in the Quebec Network of Genetic Medicine. J Inher Metab Dis 11:45–55.
19. Scriver C, Gregory D, Bernstein M, et al. 1987. Feasibility of chemical screening of urine for neuroblastoma case finding in infancy in Quebec. Can Med Assoc J 136:952–956.

20. Auray-Blais C, Giguere R, Lemieux B. 1989. Thin-layer chromatography of urinary homovanillic acid and vanillylmandelic acid for large-scale neuroblastoma mass screening. Med Pediatr Oncol 17:364–367.
21. Tuchman M, Crippin PJ, Krivit W. 1983. Capillary gas-chromatographic determination of urinary homovanillic acid and vanillylmandelic acid. Clin Chem 29:828–831.
22. Davis S, Rogers MAM, Pendergrass TW. 1987. The incidence and epidemiologic characteristics of neuroblastoma in the United States. Am J Epidemiol 126:1063–1074.
23. Monobe Y, Takeuchi LA, Hachitanda Y, et al. 1995. Pathology review for neuroblastoma mass screening program in North America: a report from the Quebec Project. Submitted for presentation, Society for Pediatric Pathology.
24 Nakagawara A, Arima-Nakagawara M, Scavarda NJ, et al. 1993. Association between high levels of expression of the TRK gene and favorable outcome in neuroblastoma. N Engl J Med 328:847–854.
25. Vogan KJ, Bernstein ML, Brisson L, et al. 1993. Absence of p53 mutations in primary neuroblastoma. Cancer Res 53:5269–5273.
26. Bessho F, Hashizume K, Nakajo T, Kamoshita S. 1991. Mass screening in Japan increased the detection of infants with neuroblastoma without a decrease in cases in older children. J Pediatr 119:237–241.
27. Hayashi Y, Hanada R, Yamamoto K. 1992. Biology of neuroblastomas in Japan found by screening. Am J Pediatr Hematol Oncol 14:342–347.
28. Nakagawara A, Zaizen Y, Ikeda K, et al. 1991. Different genomic and metabolic patterns between mass screening-positive and mass screening-negative later-presenting neuroblastomas. Cancer 68:2037–2044.
29. Hachitanda Y, Ishimoto K, Hata J, Shimada H. 1994. One hundred neuroblastomas detected through a mass screening system in Japan. Cancer 74:3223–3226.
30. Ishimoto K, Kiyokawa N, Fujita H, et al. 1990. Problems of mass screening for neuroblastoma: analysis of false-negative cases. J Pediatr Surg 25:398–401.
31. Suita S, Zaizen Y, Yano H, et al. 1994. How to deal with advanced cases of neuroblastoma detected by mass screening: a report from the Pediatric Oncology Study Group of the Kyushu area of Japan. J Pediatr Surg 29:599–603.
32. Nishi M, Miyake H, Takeda T, et al. 1994. A trial to discriminate spontaneous regression from non-regression cases during mass screening for neuroblastoma. Jpn J Clin Oncol 24:247–251.
33. Hanawa Y, Sawada T, Tsunoda A. 1990. Decrease in childhood neuroblastoma death in Japan. Med Pediatr Oncol 18:472–475.
34. Hanawa Y. 1993. The recent trend of childhood neuroblastoma death in Japan. Presented at the Third International Symposium on Neuroblastoma Screening, Kyoto, Japan.
35. Cole M, Parker L, Craft A. 1993. Decrease in childhood neuroblastoma death in Japan (letter to the editor). Med Pediatr Oncol 20:84–85.
36. Huddart SN, Muir KR, Parkes S, et al. 1993. Neuroblastoma: a 32-year population-based study — implications for screening. Med Pediatr Oncol 21:96–102.
37. Sankila R, Hakama M. 1992. Survival trends for neuroblastoma patients in Finland: negative reflections on screening. Eur J Cancer 29A:122–123.
38. Parker L, Craft AW, Dale G, et al. 1992. Screening for neuroblastoma in the north of England. Br J Med 305:1260–1263.
39. Mathieu P, Favrot M, Frappaz D, et al. 1993. Le neuroblastome de l'enfant: aspects cliniques et biologiques. Une experience de depistage en France. Ann Biol Clin 51:665–688.
40. Schilling FH, Erttmann, Ambros PF, et al. 1994. Neuroblastoma with unfavourable prognostic parameters detected by mass screening: report of the German pilot study. Proc Am Soc Clin Oncol 13:421 (#1440).
41. Tuchman M, Fisher EJ, Heisel MA, Woods WG. 1989. Feasibility study for neonatal neuroblastoma screening in the United States. Med Pediatr Oncol 17:258–264.
42. Woods WG, Lemieux B, Tuchman M. 1992. Neuroblastoma represents distinct clinical-biologic entities: a review and perspective from the Quebec Neuroblastoma Screening Project. Pediatrics 89:114–118.

43. Esteve J. 1993. Some remarks on power calculation for neuroblastoma screening. Presented at the Third International Symposium on Neuroblastoma Screening, Kyoto, Japan.
44. Murphy SB, Cohn SL, Craft AW, et al. 1991. Do children benefit from mass screening for neuroblastoma? Consensus statement from the American Cancer Society Workshop on Neuroblastoma Screening. Lancet 337:344–346.

12. Screening for cancer in high-risk families

William D. Foulkes and Steven A. Narod

1. Introduction

There has been much recent progress in our understanding of the genetic basis of cancer. Although the majority of cases of cancer do not appear to be hereditary, several genetic syndromes have been identified that are characterized by an increased familial risk of cancer. The predisposing genes for many of these syndromes have now been identified as well.

There are two fundamental steps in the screening of families at high risk of cancer. In the first step, families and individuals at increased risk of cancer are identified by their clinical features and by molecular testing. A basic family history should be taken from all individuals with cancer, because almost all types of cancer may be featured in one or more of the hereditary cancer syndromes. The second step involves screening for cancer in individuals found to be at increased genetic risk. Some of the recommended screening tests (e.g., mammography) are conventional; others (e.g., pentagastrin-stimulated calcitonin levels in multiple endocrine neoplasia type 2A (MEN2A)) are specific for persons with genetic conditions. It is important that both steps in this process are critically evaluated before genetic testing is introduced.

2. General aspects of identifying susceptible individuals

Those who are predisposed to develop cancer can be identified by studying the pattern of cancers and their inheritance in a family and then specifically looking for genetic evidence that the potentially at-risk person is, or is not, at increased risk of a particular cancer. Thus the approach is both clinical and molecular.

2.1. Approaches to clinical diagnosis

Members of a cancer family may be affected with a single type of cancer, but more commonly a patient will report cancers in relatives at several sites. For example, the hereditary nonpolyposis colorectal cancer (HNPCC) syndrome

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

includes both colon and endometrial cancer, and relatives of women with familial breast or ovarian cancer are at risk for tumors at either site. For most cancer syndromes (e.g., the Li–Fraumeni syndrome (childhood sarcomas and adrenocortical cancers, early-onset breast cancers, brain tumors, and leukemias) and the hereditary breast–ovarian cancer syndrome (ovarian cancer and early-onset breast cancer)), there are no specific features of the patient or of the tumor that allow carriers of the susceptibility gene to be identified. These diagnoses are typically made when the patient, or an astute physician, is struck by an unusual excess of cancer in the family.

In some cases, all first-degree relatives of an affected individual may be considered to be at high risk for screening purposes; for example, in some centers, ultrasound screening is offered to all female first-degree relatives of patients with ovarian cancer. More commonly (e.g., for colon and breast cancer), the diagnosis of multiple affected family members, or of relatives affected at an unusually early age, is required before a patient is considered to be a candidate for intensified screening.

For cancers of several rare sites, the likelihood of an underlying genetic predisposition is particularly high. About 5% of childhood cancers appear to be hereditary [1]; retinoblastoma and adrenocortical cancer are the subtypes most frequently hereditary. Roughly 50% of children with retinoblastoma, including all those with bilateral disease, carry mutations in the *Rb* gene on chromosome 13q14. It is important that the family members of these cases be screened by ophthalmoscopy and by DNA analysis, because vision can often be preserved when tumors are diagnosed early [2].

A large proportion of childhood adrenocortical cancers are associated with germ-line mutations in the *p53* gene [3]. There are currently no established guidelines for screening children with *p53* mutations.

Hereditary factors are equally if not more important in common adult cancers. For certain sites of cancer, the possibility of a familial syndrome is sufficiently high that a detailed genetic inquiry is always warranted. Medullary thyroid carcinoma is hereditary, either as site-specific thyroid cancer or as a part of MEN2A, in 25% of patients with the disease [4]. Similarily, 23% (19 of 82) of unselected patients with pheochromocytomas in a recent study were found to be carriers of familial disorders, including 16 cases of von Hippel Lindau disease and three cases of MEN2A [5]. Cancers of the small intestine and fallopian tube are both rare, but are features of HNPCC and the breast–ovarian cancer syndromes, respectively, and their presence should signal to the physician the possibility of a genetic diagnosis.

These cancers are exceptional; for the more common cancers of adulthood, the hereditary fraction is below 10%. Among the common cancers, ovarian cancer is probably associated with the highest hereditary component. Between 3% and 7% of unselected women with ovarian cancer appear to be from families with hereditary predisposition [6,7]. The hereditary fraction of breast cancer is probably between 2% and 5% [8]. This proportion is higher for women with breast cancer diagnosed at an early age or for women with

bilateral disease. The excess familial risk for colon cancer appears to be limited to relatives of patients diagnosed before the age of 70 [9].

Multiple primary cancer diagnosed in the same individual may be due to chance, due to a common environmental exposure, the late effect of treatment of the first malignancy, or due to genetic predisposition. A high proportion of children with both retinoblastoma and osteosarcoma carry mutations of *RB-1*. Similarly, survivors of childhood adrenocortical cancer are at a high risk of developing a second childhood cancer [10]. It will be of interest to determine what proportion of women with multiple primary cancer of the colon and endometrium are carriers of mutations in *hMSH2* or *hMLH1*, the two known genes that predispose to HNPCC. Similarly, a large proportion of women with both breast and ovarian cancer may carry *BRCA1* mutations.

In general, hereditary tumors cannot be distinguished from nonhereditary ones, and the diagnosis of familial cancer is made on historical grounds. However, one or more clues may alert the physician to the possibility that a cancer syndrome is present. Breast cancer appears at an much earlier age than expected in both the hereditary breast–ovarian cancer [11] and Li–Fraumeni syndromes [12]. Hereditary colon and endometrial cancers are also of earlier onset than nonfamilial cases [13], but the ages of onset of hereditary ovarian cancers are not remarkable [14]. The pathology of hereditary cancers may be different from their sporadic counterparts. Breast cancers associated with *BRCA1* are characterized by an increased frequency of medullary histology, and greater tumor cell proliferation rates than nonhereditary cancers [15]. *BRCA2* is associated with an excess of tubulolobular cancers [15]. Mucinous ovarian cancers are underrepresented in *BRCA1* carriers [16].

Some tumor types may be associated with preneoplastic lesions in adjacent tissues or with other characteristic benign features. Multiple polyposis associated with colon cancer is typical of FAP. The presence of generalized C-cell hyperplasia in a patient with a medullary thyroid carcinoma is a signal that the patient is from a family with MEN2A. There is no convincing evidence yet that a preneoplastic lesion is associated with hereditary breast cancer due to *BRCA1* and *BRCA2*, although atypical hyperplasia appears to be more frequent in families at increased risk [17]. In Cowden disease, hyperproliferation of breast tissue is a feature, although it is not certain whether this is a marker or a precursor of cancer (M. Peacocke, personal communication). For a few well-defined genetic syndromes, the physical appearance of the patient may be used presumptively to identify carriers of mutant alleles before cancer develops. Included in this category are neurofibromas and cafe-au-lait spots in neurofibromatosis type 1 (NF1), posterior cataracts in neurofibromatosis type 2 (NF2), and congenital hyperpigmentation of the retina in FAP. Retinal angiomas are suggestive but not diagnostic of von Hippel–Lindau syndrome. Similarly, the presence of multiple dysplastic nevi may point towards the dysplastic nevus/atypical mole syndrome. The presence of more than two trichilemmomas (benign tumors of the outer hair root sheath) is regarded as diagnostic of Cowden disease [18]. These may be difficult to diagnose without

an excision biopsy. Other skin manifestations of Cowden disease may be subtle and difficult to distinguish from lesions seen in the general population [18].

2.2. Approaches to molecular diagnosis

For most genetic cancer syndromes, it is now possible to employ specific DNA tests to identify family members who carry predisposing mutations. One of the goals of genetic counseling in oncology is to identify carriers of predisposing genetic mutations prior to the development of clinical symptoms of cancer. It is hoped that presymptomatic detection of carriers using DNA testing will permit screening efforts to concentrate on those individuals at highest risk.

There are currently two approaches to molecular diagnosis. If the sequence of a susceptibility gene is known in its normal and mutant forms, then a mutation can be sought directly in the DNA taken from lymphocytes in the blood. For many cancer syndromes, a predisposing gene has been identified and mutations have been characterized, including hereditary retinoblastoma, familial adenomatous polyposis (FAP), the Li–Fraumeni syndrome, NF1, NF2, von Hippel–Lindau disease, HNPCC, and the breast–ovarian cancer syndrome. The majority of the cancer syndromes appear to be due to mutations of tumor suppressor genes. This designation implies that both copies of the gene are rendered inactive in the tumor cells, the first by inheritance, the second by somatic mutation. Under this model, mutations that lead to premature stop codons, and thereby prevent the synthesis of a functional protein, should have a similar effect at the cellular level. It is not surprising, therefore, that mutations in genes responsible for retinoblastoma, for FAP, for NF2, and for the breast–ovarian cancer syndrome have been found dispersed over many exons. Sequencing these large genes is a formidable task, even if intermediate steps of mutation screening such as SSC (single-strand conformation) analysis are taken, and diagnostic screening tests for these diseases are not yet routine.

The introduction of various types of protein truncation test (PTT) may permit a more efficient sequential mutation analysis of genes implicated in inherited cancer syndromes where chain-terminating mutations are common, such as *APC*, *BRCA1*, and *NF2* [19]. The problem in genetically heterogenous diseases is that other genes may be responsible, and therefore even a negative result is not sufficient evidence to rule out the possibility of inheritance of a mutation in a disease-associated gene. This is where linkage analysis can be extremely valuable. Using linkage analysis, it may be possible to infer the presence of a predisposing mutation by looking for cosegregation of cancer with a particular chromosome marker. This approach requires that multiple affected family members be available for sampling. The technique also requires that the chromosomal position of a cancer susceptibility gene is known. Even if several mutations of a gene predispose to disease, a single set of DNA markers may be used. If a genetic marker exists in the population in two or

more distinguishable forms (alleles), it is said to be polymorphic and is potentially useful for linkage. The variation in DNA in human populations may be due to single base substitutions or to differences in the number of copies of a short, repeated sequence. If an affected parent carries distinguishable marker alleles on his or her two homologous chromosomes, then his or her children will be candidates for screening by linkage. If a child inherits the marker allele on the chromosome carrying the mutation, the child will likely have inherited the marker as well. However, the marker and the mutation may be separated by a meiotic crossover in the germ cell of the affected parent. When this occurs, a child with the allele of risk may be falsely predicted to be a carrier. In practice, many ordered markers on both sides of the susceptibility gene are used to construct a haplotype for each individual in a family, and the haplotype associated with the cancer mutation is identified by inspection.

Linkage analyses are technically simple, but the data may be very difficult to interpret, especially for complex diseases. In linkage analysis, the presence of a mutation is not observed but is inferred from the patterns of chromosomal segregation. Erroneous inferences will have serious consequences to the individuals receiving counseling, and linkage data must be interpreted with caution, especially for genetically heterogeneous diseases. The diagnoses of cancer must be confirmed because misclassification of one individual may lead to erroneous predictions for many. The false-positive rates in the offspring are directly related to the recombination fractions and the disease gene.

Predictive testing by linkage or by direct sequencing is greatly simplified if the disease is genetically homogenous (i.e., if all affected families are due to mutations of a single gene). Currently, this appears to be true for retinoblastoma, for FAP, and for the neurofibromatoses, but the possibility that an unlinked family may be identified in the future still remains. For other diseases, e.g., hereditary breast–ovarian cancer and HNPCC, more than one susceptibility gene appears to be involved. Failure to identify a predisposing mutation may be due to incomplete test sensitivity or to genetic heterogeneity — but it may also be that the observed "cancer family" is due to chance.

In summary, it is now possible to perform predictive testing for several hereditary cancer syndromes. Direct DNA testing is preferable, but may be limited by the size of the gene to be screened and the sensitivity of the screening method. The techniques of linkage analysis are straightforward in comparison, but the interpretation of linkage data is difficult, and should be attempted only with due attention to all possible sources of uncertainty.

3. Clinical and molecular features of selected hereditary cancer syndromes

3.1. The breast–ovarian cancer syndrome

About 4% of breast cancer cases and about 6% of ovarian cancers are believed to be hereditary [6,7]. Detecting those at risk depends upon inspection

of the family history and subsequent molecular investigation (linkage, mutation analysis). There are known to be at least two genes responsible for the breast–ovarian cancer syndrome, namely, *BRCA1* and *BRCA2*. By evaluating linkage and pedigree data, it was found that about 75% families with three or more cases of breast cancer diagnosed before age 60 and at least one case of ovarian cancer are linked to *BRCA1*. If two ovarian cancers but no male breast cancer cases are present in the pedigree, then the prior probability of linkage rises to 92% [20]. Those with mutations may also be at increased risk of prostate and colon cancer. On average, 85% of women with a *BRCA1* mutation will develop breast cancer, and 45% will develop ovarian cancer by age 70 [21]. About one third of site-specific hereditary breast cancer is due to *BRCA1*, and two thirds is due to *BRCA2*.

BRCA2 has now been cloned [21a]. From examination of pedigrees and preliminary linkage data, it appears that *BRCA2* is responsible for most of the breast-cancer-only families, particularly when male breast cancer is present. Ovarian cancer has been reported in some *BRCA2* pedigrees. Other cancers, such as prostate, colon, and laryngeal cancer, are also probably in excess [22].

BRCA1 mutations are divided into three main types: (1) frameshift, nonsense, and microdeletions resulting in a truncated protein product; (2) missense mutations, and (3) mutations that alter the stability of the messenger RNA such that only the wild-type allele is detectable in cDNA isolated from affected individuals. Because of these three main types, only a stepwise comprehensive mutation analysis approach is likely to detect most of the mutations. Using the approach of direct sequencing, conformation analysis, and cDNA allele exclusion analysis, it has been possible to detect 16 out of 20 mutations in one series of *BRCA1*-linked pedigrees [23].

Some *BRCA1* mutations are recurrent. A mutation 185delAG (a deletion of two base pairs (A and G) at codon 185) is common and appears to be mainly limited to Ashkenazi Jews [24]. This mutation may be present in 1% of unselected Ashkenazi Jews [25]. Another common mutation is 5382insC, which has been seen in Jews, Italians, and Russians and is probably an ancient mutation (unpublished observations).

3.2. Familial adenomatous polyposis and other intestinal polyposis syndromes

Familial adenomatous polyposis (FAP) is the underlying cause of about 1% of all colorectal cancer. The trait is autosomal dominant. The incidence of colorectal cancer approaches 100% in affected persons who are not treated by preventive colectomy [26].

The severity of the syndrome varies both within and between families and cannot be entirely explained by the position of the mutation in the causative gene, APC. Detection of mutaions in APC is difficult because of the large size of the gene. However, over 90% of all mutations result in a truncated protein product, so protein truncation assays have been employed and can, at best,

detect over 80% of mutation in affected individuals. This is a considerable improvement on conformation-based assays [19].

Other rare forms of hereditary polyposis include Peutz–Jeghers and juvenile polyposis. In juvenile polyposis, the cumulative incidence of intestinal cancer is less than in FAP and may not justify a similar preventive surgical approach [27]. In Peutz—Jeghers syndrome, there is an excess of death from both intestinal and extraintestinal cancers. The relative risk of cancer in those affected is approximately 20-fold, and this may justify special surveillance in these families [28]. The genes that cause these inherited syndromes have not yet been localized.

3.3. Hereditary nonpolyposis colorectal cancer

HNPCC families show an excess of early-onset colorectal cancer. This occurs in association with other cancers, either in the individual (multiple primary tumors) or in other members of the family [29]. The large intestine of those affected by colorectal cancer does not appear to have an excessive number of polyps [30]. Although right-sided colon cancers in the presence of a few scattered polyps is the classical picture [29], rectal cancer is also seen in the disease, and pedigrees may contain affected persons with either no polyps or many polyps. The polyps tend to be tubulovillous or villous rather than tubular [30]. The mean age of diagnosis of colorectal cancer in HNPCC is in the range 40–45 years [31], compared with 40 years in familial adenomatous polyposis [32] and approximately 68 years in the general North American population [33]. Other cancers that are associated with the syndrome include adenocarcinomas of the endometrium, stomach, small intestine, and ovary. Transitional cell carcinomas of the upper urothelium are also seen [29].

DNA from several forms of hereditary cancer associated with HNPCC is prone to erroneous replication during mitosis. This leads to the generation of new fragment sizes in repeat sequences. If these altered sequences occur near or within genes, this could disrupt their function. This phenomenon of microsatellite instability (the RER+ phenotype) is present in approximately 15% of unselected sporadic colon cancers [34], 58% of colon cancers occuring in those less than 35 years of age [35], and 0 up to 3% of sporadic adenomas [36]. In patients from families with HNPCC, 86% of colon cancers and 57% of adenomas showed instability [36]. Microsatellite instability is particularly common in right-sided colon cancers, irrespective of family history, and is also seen in apparently sporadic bladder, lung, head and neck, germ cell, and other cancers [37]. It remains to be seen if this assay will be useful in identifying family members at significantly increased risk.

Finding mutations in the HNPCC genes has so far been most successfully achieved by using an in vitro translation technique first used for detecting abnormal APC transcripts [35]. Others using SSCA have not found deletions or nonsense mutations to be as frequent as missense mutations [38]. However,

relatively few mutations have been reported, and routine genetic testing is not widely available.

3.4. Neurofibromatosis type 1

Von Recklinghausen's neurofibromatosis (NF1) is one of the most common human genetic disorders, affecting 1 in 3000 people. There is a very high proportion of new mutations — around 50%. This means it is common for an affected person to have no relatives with NF1, but nevertheless their offspring have a 50% risk of being affected. The main clinical features are cafe-au-lait spots and subcutaneous and cutaneous neurofibromas (seen in more than 90% of gene carriers). Death can result from neurofibrosarcomas, optic nerve gliomas, and pheochromocytomas. These tumors collectively affect about 5% of gene carriers. Variable expressivity is the rule and very mildly affected individuals may have severely affected offspring.

Finding mutations in the huge NF1 gene (mRNA 13kb) is difficult. As of January 1994, only 45 mutations had been found in 500 affected individuals. Because of the relatively benign course of neurofibromatosis type 1 in many affected individuals, and the unpredictability of expression, there has not been the same urgency for mutation detection as has been seen for other cancer-susceptibility syndromes [39].

3.5. Neurofibromatosis type 2

Unlike NF1, neurofibromatosis type 2 (NF2) is often a progressive disorder leading to early death. The principal features are acoustic schwannomas (which may be bilateral), meningiomas and astrocytomas, and other tumors arising from the central or peripheral nervous system. The onset of symptoms varies widely, but potentially fatal tumors can develop in children and adolescents.

In many situations, DNA analysis will permit the majority of the individuals in the family who are believed to be at risk to be excluded from further investigation. For example, 82 individuals in 13 families with neurofibromatosis type 2 (NF2) were considered to be at risk for developing acoustic neuromas and meningiomas prior to genetic testing [40]. By using chromosome-22 linked markers, it was possible to effectively rule out the NF2 carrier state for 65% of these. The 13 individuals (16%) who were found to be at high risk for the disease could be offered more intensive screening. When the NF2 gene was cloned, it became possible to do direct screening of DNA for mutations. Because of the relatively large size of the NF2 gene, DNA is often screened for abnormal banding patterns with single-strand conformation analysis (SSC) before sequencing is done. Using this two-step process, it was possible to identify mutations in 60% of 58 unrelated individuals with NF2 (M Ruttledge, personal communication).

3.6. Type 2 multiple endocrine neoplasia

The multiple endocrine neoplasias type 2 (MEN2) consists of three overlapping syndromes: MEN2A, MEN2B, and familial medullary carcinoma of the thyroid (FMTC). MEN2A features medullary C-cell hyperplasia leading to medullary thyroid cancer (MTC) seen in association with hyperparathyroidism and pheochromocytoma. MEN2B is similar, with the addition of mucosal neuromas and other physical signs. In FMTC, only the thyroid is affected. One gene, *RET*, is responsible for all three syndromes [41].

Direct sequencing is particularly useful because mutations in a small number of codons are responsible for the majority of MEN2A families. Over 95% of the mutations in 118 families with this syndrome were found at only five cysteine residues in the extracellular domain of the *RET* oncogene, and direct testing for this syndrome has become straightforward within one year of the identification of the underlying gene defect [42].

Up to 10% of cases of FMTC without a family history may have germ line mutations in *RET* [43] (G. Lenoir, personal communication); because of the potential for cure with early surgery and the lack of early symptoms (see section 4.5), it is prudent to offer genetic testing to all relatives of a sporadic case of FMTC.

3.7. Von Hippel–Lindau syndrome

The von Hippel–Lindau (VHL) syndrome is quite rare, affecting about 1 in 35,000 births. It is characterized by the occurrence of tumors at multiple sites, including kidney, cerebellum, eye, pancreas, adrenal gland, and epididymis. Early-onset clear cell carcinoma of the kidney occurs in about 40% of gene carriers, and pheochromocytoma in 18% [44] (19% of those with sporadic pheochromocytoma have mutations in the VHL gene [5]).

The causative gene contains only three exons and encodes a mRNA of 0.8 kb, with a predicted protein of 284 amino acids. Mutations have been detected in 75% of 114 VHL families. Unlike the other cancer syndromes described above, most of the mutations are nucleotide substitutions. There is a suggestion from genotype–phenotype correlations that pheochromocytomas are uncommon, if not entirely absent, in those with inactivating mutations of VHL [45].

4. Prevention of and screening for cancer in predisposed individuals

Individuals at high risk of cancer as a result of hereditary mutations in cancer susceptibility genes require careful management. It is in these high-risk groups that prevention and early detection programs may be shown to be most effective. For several cancer syndromes, the relative merits of prophylactic

surgery and intensified screening must be considered. For example, young adults found to carry a gene mutation in the *RET* oncogene, which predisposes to medullary cancer of the thyroid in the context of MEN2A, may be followed by annual screening using the pentagastrin stimulation test [46] or may opt for prophylactic thyroidectomy. Since the risks and benefits differ for different cancer syndromes, we discuss each clinical problem in turn. In all cases, it is important to evaluate the risk of death from cancer in the absence of intervention, with screening, with preventive medical intervention, with surgery or with a combination of these options. The risk of death or disability from the procedures must also be considered. Cancer may still occur after preventive surgery because of pre-existing microscopic foci of tumor that have metastasized at time of surgery or because of residual tissue remaining at risk.

4.1. Breast–ovarian cancer syndromes

A person who carries a mutation in *BRCA1* has an 85% risk of breast cancer by age 70 [21]. Therefore, prevention and early detection of familial breast cancer is of immediate concern. There is little evidence that screening of young women by mammography will reduce the mortality from breast cancer (see chapter 6, this volume). It has not yet been possible to address this question specifically in women at high familial risk, but there is no reason to believe that sensitivity of the mammography is better or that the natural history of breast cancer is different for women at increased risk because of family history.

Chemoprophylaxis and preventive surgery are potential alternatives to screening. The most obvious candidate for chemoprevention is tamoxifen, which is now under evaluation in above-average-risk women in an NSABP trial. Its use to treat breast cancer resulted in a 39% reduction in the occurrence of contralateral breast cancer [47]. Tamoxifen has a number of side effects, some of which are serious. There is an increase in the lifetime risk of endometrial cancer and a possible increased risk of colorectal cancer [48].

An alternative to chemoprevention is surgery. Prophylactic total mastectomy and prophylactic subcutaneous mastectomy have been carried out for many years [49–51]. Despite the extensive literature pertaining to these operations, very little is known about the effectiveness of these procedures. However, even a lifetime risk of breast cancer of 5% following surgery might be acceptable to a *BRCA1* mutation-carrying woman in her late 30s, approaching the peak incidence of breast cancer. This would represent a reduction in risk of at least 10-fold compared with no intervention [21].

Carrriers of *BRCA1* mutations also have approximately a 50% risk of ovarian cancer to age 70 [21]. There is currently much interest in evaluating ovarian ultrasound in screening for ovarian cancer. Transabdominal [52] and transvaginal [53] ultrasound (with or without color-flow Doppler) and serum CA125 screening tests [54], either alone or in combination [55], have all been assessed as potential screening tests in various settings. However, no single test has yet proved to be practical for population screening, for several reasons: (1)

ovarian cancer is relatively uncommon (less than 2000 cases per year in Canada); (2) there are no easily identifiable premalignant stages; and (3) there is no single screening procedure that has a sufficiently large positive predictive value. This implies that a large number of women have to undergo further investigation in order to detect one cancer.

Women at increased risk of ovarian cancer as a result of their family history and/or *BRCA1* test are a subgroup likely to be suitable for screening. Using transvaginal ultrasound and color-flow Doppler monitoring, 61 abnormal lesions were detected in 1601 self-referred women with family history of ovarian, breast, and other cancers. These 61 women were referred for surgical investigation, and six were found to have cancers, five of which were early stage. It should be noted, however, that only one of the six women had a family history that was strongly suggestive of hereditary breast and ovarian cancer [56]. Using transvaginal ultrasound with color flow imaging, Weiner et al. [57] found ovarian cancer in 4 of 600 women with a past history of breast cancer (this group is at roughly double the general public's risk of ovarian cancer). Muto et al. [58] screened 386 women with a family history of ovarian cancer using transvaginal sonography, color-flow Doppler, and CA-125. The ultrasound examination was abnormal in 23% of the women, but no malignant ovarian lesion was detected.

Screening with multiple serum markers has been proposed in an attempt to improve detection of early ovarian cancers. At least one of three markers *(M-CSF*, *OVX1*, and *CA125)* were elevated in the serum of 45 out of 46 women with stage I ovarian carcinomas. Unfortunately 51% of women with benign pelvic masses also had elevations of one or more markers [59].

The benefits and limitations of ovarian cancer screening should be discussed with the woman and compared with prophylactic oophorectomy. Because of the high lifetime risk of ovarian cancer associated with *BRCA1*, and because the sensitivity and effectiveness of the current methods of screening are uncertain, several groups currently recommend prophylactic removal of the ovaries of *BRCA1* mutation carriers around the time of menopause [60,61]. Unfortunately, about 5% of these women will later develop peritoneal cancer [62,63].

4.2. Familial adenomatous polyposis

Until the FAP gene, *APC*, was cloned, the diagnosis of FAP was based on lower bowel surveillance, ophthalmic examination for congenital hypertrophy of the retinal pigment epithelium, and skin examination for epithelioid cysts. Linkage and direct mutation analysis are now possible. Because of the difficulty in identifying mutations, all three approaches are in use [64].

The high mortality associated with FAP has been noted since the syndrome was first recognized. Prophylactic surgery has been used for FAP since 1934. Because the risk of colorectal cancer is almost 100% in FAP [32], surgery is recommended to those who have inherited a disease-associated allele of FAP.

The risk of dying from colorectal cancer is greatly reduced by appropriate surgery, followed where necessary by lower bowel surveillance. After colorectal cancer, the most common cause of death in pedigrees with FAP is upper gastrointestinal cancer. Gastroduodenal polyposis is common in FAP, affecting at least 45% of gene carriers [65]. Management is problematic. Regular but infrequent video endoscopy has been assessed [66], and sulindac chemoprophylaxis has been advocated. However, invasive cancer can be difficult to detect if the duodenum is carpeted with polyps.

Desmoid tumors are an important cause of morbidity and mortality in FAP. A rare feature of FAP is brain tumors. The occurrence of polyposis and brain tumor is referred to as Turcot syndrome. It now appears that most cases of Turcot syndrome are due to mutations in *APC*. Some of the remaining cases are caused by mutations in the HNPCC genes. Early detection by MRI has been advocated, but its effectiveness in preventing death is completely unknown [67].

4.3. Hereditary nonpolyposis colorectal cancer

While it is currently accepted that preventive colectomy reduces mortality for adults with FAP, the relative value of screening, versus prophylactic surgery, for carriers of genes for HNPCC has not yet been established. There is some evidence that, in the general population, screening sigmoidoscopy may be beneficial in reducing mortality from cancer of the rectum and the distal colon (see chapter 5, this volume) and that colonoscopy, followed by polypectomy when indicated, reduces the incidence of invasive colon cancer [68]. Colonoscopy is currently recommended from an early age in HNPCC families. However, it should be noted that the natural history of hereditary colon cancer may differ from sporadic cancer. Hereditary colon cancer is usually right-sided and beyond the reach of the sigmoidoscope [29]. Furthermore, it has not been established whether colon cancers in HNPCC families occur in pre-existing polyps. Even if all HNPCC colorectal cancers do arise from polyps, the rate at which the adenoma–carcinoma sequence progresses may be so fast as to limit the effectiveness of yearly colonoscopy. Six individuals from the extensively characterized Dutch pedigrees were each found to have an advanced right-sided colon cancer between screening tests (colonoscopy in five and barium enema in one). Three of the tumours were Dukes' stage C, and they occurred only 26–31 months after a previous screening [69]. Lynch reports that one woman with endometrial cancer diagnosed at age 36 (as part of HNPCC) later presented with a transverse colon carcinoma 18 months after one of her (two-yearly) colonoscopies. Reducing the screening interval to six months did not prevent cancer: five months after a negative colonoscopy, she was found to have synchronous carcinomas of the caecum and rectum (Henry Lynch, personal communication). While these may be unusual examples, it will be of primary importance to establish the relative benefits of screening versus prophylactic surgery in individuals who carry HNPCC gene mutations.

It is important to realize that many organs are at risk in HNPCC, and it will be impossible to prevent or screen for all possible cancers. The cumulative incidence of endometrial cancer in HNPCC in gene carriers to age 70 is 30%, 10 times the expected figure. The peak incidence of the disease occurs at least 15 years earlier in gene carriers than in the general population [70]. The risk of ovarian cancer in HNPCC is mildly elevated. Breast cancer does not seem to be a part of HNPCC [13,71].

4.4. Prostate cancer

The value of population screening for prostate cancer is still under debate (see chapter 7, this volume). However, because of an increased likelihood of cancer in high-risk individuals, screening is probably justified in some subgroups. Recent studies have demonstrated that a family history of prostate cancer is a risk factor for prostate cancer [72,73]. Segregation analysis suggests that about 10% to 12% of all prostate cancer is due to an inherited gene [72]. The nature of inheritance is not certain. The risk is probably highest in brothers of cases diagnosed at a young age [74], although this may be partly accounted for by recall bias falsely reducing the risk attributable to fathers [73]. In one study, 10% of men who had brothers with prostate cancer had prostate cancer identified when investigated [73]. Family history was only significantly predictive of prostate cancer in those with a negative digital rectal examination. This is probably because of a high false-positive rate in those with a family history, possibly due to a higher incidence of benign prostatic hypertrophy in relatives of prostate cancer cases.

Because the prevalence of cancer among genetically predisposed individuals is usually higher than among the general public, it is expected that the positive predictive value of the screening test employed will be greater for high-risk individuals than for unselected subjects. For example, a PSA level of greater than 3.0ng/ml was associated with a positive predictive value for prostate cancer of 18.9% overall, but was 28.6% among men with a positive family history of prostate cancer [73]. The positive predictive value of a screening test is particularly important when the confirmatory testing carries the potential for surgery-related morbidity.

4.5. Multiple endocrine neoplasia

It possible to manage those with known *RET* mutations by regular pentagastrin challenge or by prophylactic thyroidectomy. The risks and benefits of these approaches can now be assessed. Screening by pentagastrin challenge was established years before the responsible gene was identified, but now that direct mutation detection is a straightforward procedure, the value of such screening should be questioned (especially since it is disliked by at-risk individuals). This is clearly shown from a large series from the Mayo Clinic [75]. Here 63 patients had total thyroidectomies for MEN disease. No patient

died as a result of the procedure, but permanent hypoparathyroidism resulted in 5%. In MEN2A and 2B, parathyroid surgery may also be necessary: the incidence of permanent hypoparathyroidism was 7% [75]. Importantly, 82% of screen-detected cases had invasive MTC and 10% had metastases. Of the nine patients who died during follow-up, three died of metastatic MTC. MTC in MEN2B may be more aggressive than in MEN2A, since all cases over the age of seven months old with MEN2B had invasive carcinoma at surgery. In a series of 13 thyroidectomy patients from seven MEN2A pedigrees, only seven patients had elevated provoked plasma calcitonin levels, but all cases had C-cell hyperplasia, with or without MTC [76]. In a small series of 16 children between the ages of 6 and 16 who had a total thyroidectomy for thyroid carcinoma, all were alive and well 1 to 21 years later [77]. Prophylactic, early surgery (in infancy or very early childhood) by an experienced pediatric endocrine surgeon is probably the management of choice in those who test positive for *RET* mutations. In order to detect pheochromocytomas, urinary vanyll mandelic acid (VMA) should be regularly monitored. Screening for cancer at other sites can be difficult.

4.6. von Hippel–Lindau disease

Mutations in *VHL* should be quite easy to detect because of the small size of the gene [44,45]. Those at risk will require extensive and regular screening. In known carriers, the most important organs in which to look for early signs of disease are the kidney and adrenal gland. Renal ultrasound is the most convenient technique. If tumors are detected, conservative, nephron-sparing surgery is essential because of the likelihood of multiple, bilateral clear cell renal carcinomas. For pheochromocytomas, families with missense mutations are more likely to have pheochromocytomas than those with inactivating mutations [45], but the correlation is not absolute, and some are not willing to exclude pheochromocytoma screening on this basis [78]. Regular blood pressure measurements and 24-hour urinary VMA levels may be used to detect what can be microscopic unilateral or bilateral tumors. However, both blood pressure and urinary catecholamine metabolites can be normal or near normal in those with proven pheochromocytomas [79]. The routine use of other investigations such as CT or MRI to detect these tumors is uncertain. MRI may be particularly useful in following the progress of hemangioblastomas [80], but whether this will improve prognosis is not yet known.

5. Conclusion

In this chapter we have described an approach to cancer screening in high-risk families. It is important to recognize phenotypic signs of cancer syndromes so that at-risk individuals can be promptly counseled. Where appropriate, screening and intervention can be planned. Now that DNA testing is becoming more

widely available, it may be possible to determine who is at high risk before any symptoms or signs are present, and this possibility may be attractive to those attempting to target screening to those at highest risk. In the future, it will be important to compare different preventive regimens. The cloning of cancer susceptibility genes will result in new insights into their function and may lead to advances in cancer drug development in which drug-based prevention of cancer may be an important part.

References

1. Narod SA, Stiller C, Lenoir GM. 1991. An estimate of the hereditary fraction of childhood cancer. Br J Cancer 6:993–999.
2. Weiss AH, Karr DJ, Kalina RE, Lindsley KL, Pendergrass TW. 1994. Visual outcomes of macular retinoblastoma after external beam radiation therapy. Ophthalmology 101:1244–1249.
3. Wagner J, Portwine C, Rabin K, Leclerc J-M, Narod SA, Malkin D. 1995. A high frequency of germline p53 mutations in childhood adrenocortical cancer. J Natl Cancer Inst 86:1707–1710.
4. Saad RK, Ordonez NG, Rashid RK, et al. 1984. Medullary carcinoma of the thyroid. A study of the clinical features and prognostic factors in 161 patients. Medicine 63:319–342.
5. Neumann HPH, Berger DP, Sigmund G, et al. 1993. Pheochromocytomas, multiple endocrine neoplasia type 2 and Von Hippel–Lindau disease. N Engl J Med 329:1531–1538.
6. Narod SA, Madlensky L, Bradley L, et al. 1994. Hereditary and familial ovarian cancer in Southern Ontario. Cancer 74:2341–2346.
7. Takahashi H, Behbakht K, McGovern PE, et al. 1995. Mutation analysis of the BRCA1 gene in ovarian cancers. Cancer Res 55:2998–3002.
8. Lynch HT, Lynch JF. 1986. Breast cancer in an oncology clinic: 328 consecutive patients. Cancer Genet Cytogenet 23:369.
9. Fuchs CS, Giovannucci EL, Colditz GA, Hunter DJ, Speizer FE, Willett WC. 1994. A prospective study of family history and the risk of colorectal cancer. N Engl J Med 331:1669–1674.
10. Hartley AL, Birch JM, Marsden HB, et al. 1987. Adrenal cortical tumors: epidemiological and familial aspects. Arch Dis Child 62:683–689.
11. Claus EB, Risch N, Thompson WD. 1990. Age of onset as an indicator of familial risk of breast cancer. Am J Epidemiol 131:961–972.
12. Garber JE, Goldstein AM, Kantor AF, Dreyfus MG, Fraumeni JF, Li FP. 1991. Follow up study of twenty-four families with the Li–Fraumeni syndrome. Cancer Res 51:6094–6097.
13. Watson P, Lynch HT. 1993. Extracolonic cancer in hereditary nonpolyposis colorectal cancer. Cancer 71:677–685.
14. Amos CI, Shaw GL, Tucker MA, Hartge P. 1992. Age at onset for familial epithelial ovarian cancer. JAMA 268:1896–1899.
15. Marcus JN, Watson P, Page DL, et al. 1996. Hereditary breast cancer: pathobiology, prognosis and BRCA1 and BRCA2 linkage. Cancer 77:697–709.
16. Narod SA, Tonin P, Lynch H, et al. 1994. Histology of BRCA1-associated ovarian tumours. Lancet 343:236.
17. Dupont WD, Page DL. 1987. Breast cancer risk associated with proliferative disease, age at first birth and a family history of breast cancer. Am J Epidemiol 125:769–779.
18. Salem OS, Steck WD. 1983. Cowden's disease (multiple hamartoma and neoplasia syndrome). Case report and review of the English literature. J Am Acad Dermatol 8:686–696.
19. Powell SM, Petersen GM, Krush AJ, et al. 1993. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 329:1982–1987.

20. Narod SA, Ford D, Devilee P, et al. 1995. An evaluation of genetic heterogeneity in 145 breast-ovarian cancer families. Am J Hum Genet 56:254–264.
21. Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE. 1994. The risks of cancer in BRCA1 mutation carriers. Lancet 343:692–695.
21a. Wooster R, Bignell G, Lancaster J, et al. 1996. Identification of the breast cancer susceptibility gene BRCA2. Nature 378:783–792.
22. Wooster R, Neuhausen S, Mangion J, et al. 1994. Localisation of a breast cancer susceptibility gene (BRCA2) to chromosome 13q by genetic linkage analysis. Science 265:2088–2090.
23. Serova O, Narod SA, Tonin P, et al. 1996. A high incidence of BRCA1 mutations in 20 breast-ovarian families. Am J Hum Genet.
24. Tonin P, Serova O. Lenoir G, et al. 1995. BRCA1 mutations in Ashkenazi Jewish women. Am J Hum Genet 57:189.
25. Struewing JP, Abeliovich D, Peretz T, Avishai N, Kaback MM, Collins FS, Brody LC. 1995. The carrier frequency of the BRCA1 185delAG mutation is approximately 1 per cent in Ashkenazi Jewish individuals. Nature Genet 11:198–200.
26. Campbell WJ, Spence RAJ, Parks TG. 1994. Familial adenomatous polyposis. Br J Surg 81:1722–33.
27. Desai DC, Neale KF, Talbot IC, Hodgson SV, Phillips RKS. 1995. Juvenile polyposis. Br J Surg 82:14–17.
28. Giardello FM, Welsh SB, Hamilton SR, et al. 1987. Increased risk of cancer in the Peutz–Jeghers syndrome. N Engl J Med 316:1511–1514.
29. Lynch HT, Smyrk TC, Watson P, et al. 1993. Genetics, natural history, tumor spectrum and pathology of hereditary non-polyposis colorectal cancer: an updated review. Gastroenterology 104:1535–1539.
30. Jass JR, Stewart SM, Stewart J, Lane MR. 1994. Hereditary non-polyposis colorectal cancer — morphologies, genes and mutations. Mutation Res 310:125–133.
31. Albano WA, Recarbaren JA, Lynch HT, et al. 1982. Natural history of hereditary cancer of the breast and colon. Cancer 50:360–363.
32. Bussey HJR. 1975. Familial Polyposis Coli: Family Studies, Histopathology, Differential Diagnosis and Results of Treatment. Baltimore: Johns Hopkins University Press.
33. Parkin DM, Muir CS, Whelan SL, Gao Y-T, Ferlay J, Powell J. 1992. Cancer Incidence in Five Continents, vol. VI (IARC scientific publication No. 120). Lyon: International Agency for Research on Cancer.
34. Lothe RA, Peltomaki P, Meling GI, et al. 1993. Genomic instability in colorectal cancer: relationship to clinicopathological variables. Cancer Res 53:5849–5852.
35. Lui B, Farrington SM, Peterson GM, et al. 1995. Genetic instability occurs in the majority of young patients with colorectal cancer. Nature Med 1:348–352.
36. Aaltonen LA, Peltomaki P, Mecklin J-P, et al. 1994. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Res 54:1645–1648.
37. Marra G, Boland CR. 1995. Hereditary nonpolyposis colorectal cancer: the syndrome, the genes, and historical perspectives. J Natl Cancer Inst 87:1114–1125.
38. Han H-Y, Maruyama M, Baba S, Park J-G, Nakamura Y. 1995. Genomic structure of mismatch repair gene, hMLH1, and its mutation analysis in patients with hereditary nonpolyposis colorectal cancer (HNPCC). Hum Mol Genet 4:237–242.
39. Upadhyaya M, Shaw DJ, Harper PS. 1994. Molecular basis of neurofibromatosis type 1: mutation analysis and polymorphisms in the NF1 gene. Hum Mutat 4:83–101.
40. Ruttledge MH, Narod SA, Dumanski JP, et al. 1993. Pre-symptomatic diagnosis for neurofibromatosis type 2 employing a combination of chromosome 22 markers. Neurology 43:1753–1760.
41. van Heyningen V. 1994. One gene — four syndromes. Nature 367:319–320.
42. Mulligan LM, Eng C, Healey CS, et al. 1994. Specific mutations of the RET proto-oncogene are related to disease phenotype in MEN2A and FMTC. Nature Genet 6:70–74.

43. Eng C, Mulligan LM. Smith DP, et al. 1995. Mutation of the RET protooncogene in sporadic medullary thyroid cancer. Genes Chromosom Cancer 12:209–212.
44. Linehan WM, Lerman MI, Zbar B. 1995. Identification of the von Hippel Lindau (VHL) gene: its role in renal cancer. JAMA 273:564–570.
45. Chen F, Kishida T, Yao M, et al. 1995. Germ-line mutations in the von-Hippel Lindau disease tumour suppressor gene: correlation with phenotype. Hum Mutat 5:66–75.
46. Gagel RF, Tashjian AH, Cummings T, et al. 1988. The clinical outcome of prospective screening for multiple endocrine neoplasia type 2a. N Engl J Med 318:478–484.
47. Early Breast Cancer Trialists Group. 1992. Systemic therapy of early breast cancer by hormonal, cytotoxic or immune therapy. Lancet 339:1–15,71–85.
48. Powles TJ, Hickish T. Tamoxifen therapy and carcinogenic risk. 1995. J Natl Cancer Inst 87:1343–1345.
49. Temple WJ, Lindsay RL, Magi E, Urbanski SJ. 1991. Technical considerations for prophylactic mastectomy in patients at high risk for breast cancer. Am J Surg 161:413–415.
50. Ziegler LD, Kroll SS. 1991. Primary breast cancer after prophylactic mastectomy. Am J Clin Oncol 14:451–454.
51. Pennisi VR, Capozzi A. 1989. Subcutaneous mastectomy data: a final statistical analysis of 1500 patients. Aesth Plast Surg 13:15–21.
52. Campbell S, Bhan V, Royston P, Whitehead MI, Collins WP. 1989. Transabdominal ultrasound screening for early ovarian cancer. Br Med J 299:1363–1367.
53. Kurjak A, Zalud I. 1992. Transvaginal colour flow Doppler in the differentiation of benign and malignant ovarian masses. In Sharp F, Mason WP, Creasman W (eds.), Ovarian Cancer 2: Biology, Diagnosis and Management. London: Chapman and Hall, pp. 249–264.
54. Cuckle HS, Wald NJ. 1990. The evaluation of screening tests for ovarian cancer. In Sharp F, Mason WP, Leake RE (eds.), Ovarian Cancer: Biological and Therapeutic Challenges. London: Chapman and Hall, pp. 229–239.
55. Jacobs I, Prys-Davies A, Bridges J, et al. 1993. Prevalence screening for ovarian cancer in postmenopausal women by CA125 measurement and ultrasonography. Br Med J 306:1030–1034.
56. Bourne TH, Campbell S, Reynolds KM, et al. 1993. Screening for early familial ovarian cancer with transvaginal ultrasound and colour flow imaging. Br Med J 306:1025–1029.
57. Weiner Z, Beck D, Shteiner M, et al. 1993. Screening for ovarian cancer in women with breast cancer and transvaginal sonography and color flow imaging. J Ultrasound Med 12:387–393.
58. Muto MG, Cramer DW, Brown DL, et al. 1993. Screening for ovarian cancer: the preliminary experience of a familial ovarian cancer center. Gynecol Oncol 51:12–20.
59. Woolas RP, Xu F-J, Jacobs IJ, et al. 1993. Elevation of multiple serum markers in patients with stage 1 ovarian cancer. J Natl Cancer Inst 85:1748–1751.
60. Kerlikowske K, Brown JS, Grady DG. 1992. Should women with familial ovarian cancer undergo prophylactic oophoretomy? Obstet Gynecol 80:700–707.
61. Lynch HT, Severin MJ, Mooney MJ, Lynch J. 1995. Insurance adjudication favoring prophylactic surgery in hereditary breast–ovarian cancer syndrome. Gynecol Oncol 57:23–26.
62. Tobacman JK, Tucker MA, Kase R, Greene M, Costa J, Fraumeni JF. 1982. Intra-abdominal carcinomatosis after prophylactic oophorectomy in ovarian cancer prone families. Lancet 2:795–797.
63. Piver MS, Jishi MF, Tsukada Y, Nava G. 1993. Primary peritoneal carcinoma after prophylactic oophorectomy in women with a family history of cancer. Cancer 71:2751–2755.
64. Burn J, Chapman P, Delhanty J, et al. 1991. The UK Northern Region genetic register for familial adenomatous polyposis coli: use of age of onset, congenital hypertrophy of the retinal pigment epithelium and DNA markers in risk calculations. J Med Genet 28:289–296.
65. Jagelman DG. 1987. Extracolonic manifestations of familial polyposis coli. Semin Surg Oncol 3:88–91.
66. Nugent KP, Spigelman AD, Williams CB, Talbot IC, Phillips RKS. 1994. Surveillance of duodenal polyps in familial adenomatous polyposis: progress report. J R Soc Med 87:704–706.

67. Coyle T. 1995. Gastrointestinal polyposis and nonpolyposis syndromes. N Engl J Med 332:1518–1519.
68. Winawer SJ, Zauber AG, Ho MN, et al. 1993. Prevention of colorectal cancer by colonosocopic polypectomy. N Engl J Med 329:1977–1981.
69. Vasen HFA, Nagengast FM, Meera Khan P. 1995. Interval cancers in hereditary non-polyposis colorectal cancer (Lynch syndrome). Lancet 345:1184–1185.
70. Watson P, Vasen HFA, Mecklin JP, Jarvinen H, Lynch HT. 1994. The risk of endometrial cancer in hereditary nonpolyposis colorectal cancer. Am J Med 94:516–520.
71. Vasen HFA, Den Hartog Jager FCA, Menko FH. 1989. Screening for hereditary non-polyposis colorectal cancer: a study of 22 kindreds in the Netherlands. Am J Med 86:278–281.
72. Carter BS, Steinberg GD, Beaty TH, Childs B, Walsh PC. 1991. Familial risk factors for prostate cancer. Cancer Surv 11:5–13.
73. Narod SA, Dupont A, Cusan L, et al. 1995. The impact of family history on early detection of prostate cancer. Nature Med 1:99–101.
74. Cannon L, Bishop DT, Skolnik, Hunt S, Lyon JL, Smart CR. 1982. Genetic epidemiology of prostate cancer in the Utah Mormon genealogy. Cancer Surv 1:47–69.
75. O'Riordain DS, O'Brien T, Weaver AL, et al. 1994. Medullary thyroid carcinoma in multiple endocrine neoplasia types 2A and 2B. Surgery 115:1017–1023.
76. Wells SA Jr, Chi DD, Toshima K, et al. 1994. Predictive DNA testing and prophylactic thyroidectomy in patients at risk for multiple endocrine neoplasia type 2A. Ann Surg 220:237–247.
77. Stael APM, Plukker JThM, Piers DA, Rouwe CW, Vermey A. 1995. Total thyroidectomy in the treatment of thyroid carcinoma in childhood. Br J Surg 82:1083–1085.
78. Green JS. 1995. Ph.D. dissertation, Memorial University, Newfoundland, Canada.
79. Richard S, Biegelman Duclos J-M, et al. 1994. Phaeochromocytoma as the first manifestation of von Hippel-Lindau disease. Surgery 116:1076–1081.
80. Filling-Katz MR, Choyke P, Patronas N, et al. 1991. Central nervous system involvement in von-Hippel Lindau disease. Neurology 41:41–46.

13. Screening in developing countries: problems and opportunities

Anthony B. Miller

1. Introduction

Cancer is rapidly becoming an important cause of morbidity and mortality in most developing countries. Indeed, given the population size and shifting demographics in the world, more cancers already occur in developing than developed countries [1].

As in developed countries, screening is superficially an attractive cancer control option for developing countries. However, screening programs should not be introduced into such countries by a slavish adoption of developed-country approaches, since the costs could easily overwhelm the health care system and result in an overall diminution of health status of the population. Further, screening should only be considered within the context of a carefully thought-through strategic plan for cancer control that takes into consideration the specific cancer situation of the country, the resources available for cancer control, and the other potential options for cancer control [2].

In many developing countries, the extent of disease at diagnosis currently is such that even a minimum investment in public and professional education could result in a substantial improvement in early diagnosis, and thus the stage of the disease at detection [3]. An essential prerequisite to screening is therefore such an educational program.

2. Early detection of cancer

If cancer can be diagnosed early in its course, treatment is generally more effective than when it is advanced. In developing countries, it is essential that the limitations as well as the benefits of early diagnosis and screening are recognized in order to avoid uncritical adoption of developed-country "high technology" but poor cost-effective approaches in this area, or of methods not achieving the needed coverage of the population. In addition, it is important to recognize that screening programs should not be introduced unless there is adequate manpower to perform the tests and facilities for diagnosis, treatment, and follow-up of individuals with abnormal test results.

A.B. Miller (ed), ADVANCES IN CANCER SCREENING.

2.1. Education for early diagnosis

Two means of early cancer detection are possible. One is educating people concerning the early signs and symptoms of the disease and encouraging them to obtain medical attention promptly when these occur. The public can be alerted to the potential significance of lumps; sores that do not heal; blood in sputum, the stool, urine, etc.; continued indigestion; or cough. People may also be trained to examine monthly their skin, breasts, and mouth for abnormal signs that could indicate cancer. Increasing the awareness of the general public to the problem of cancer and the greater potential for cure when it is detected early may thus promote early diagnosis. There is evidence that such approaches, combined with the availability of effective treatment, resulted in improvement in stage at presentation and reduction in mortality from cancer of the cervix in developed countries in the first half of this century [4]. Thus it is reasonable to assume that in the developing countries, where a high proportion of cancers that are relatively curable in developed countries occur at advanced stages, increased public awareness and greater physician and allied health care worker awareness, combined with prompt and effective therapy, could have a major impact on the disease.

This approach requires raising the consciousness of cancer among primary health care workers. Systematic training of such workers for the detection of some sites of cancer could result in an important benefit in developing countries without investment in high technology. This training requires determining the stage of disease of such cancers, e.g., cervix, breast, mouth, and skin, and if the majority are advanced (i.e., in stage III or IV), promoting measures for earlier diagnosis and referral. Even for cases where the eventual outcome is unchanged, treatment is simpler and the quality of life improved.

Among the measures to be considered for early diagnosis are

- Encourage medical and allied health professions to be aware of the symptoms and signs of cancer.
- Introduce public awareness campaigns concerning the symptoms of cancer and the benefits of early diagnosis.
- Where appropriate, introduce measures to encourage "downstaging" for common cancers that are potentially curable if found early.

2.2. Downstaging for cancer

Programs of visual inspection of the cervix using a speculum by specially trained health workers are being evaluated in India and other parts of Asia and in Africa. It is important to recognize that visual inspection has always been a part of cervical cytology screening, though its contribution to the early detection of cervical cancer in successful programs cannot be determined.

Downstaging for cancer of the cervix has been defined as "The detection of the disease in an earlier stage when still curable, by nurses and other nonmedical health workers using a simple speculum for visual inspection of the cervix"

[5]. Downstaging of cancer of the cervix is currently an experimental procedure that is not known to be beneficial in controlling the disease. However, for countries that have no possibility of introducing cytologic screening for cancer of the cervix for many years, downstaging should be considered in the context of general approaches to early detection of a critically important cancer. Downstaging is intended to make use of available health care resources in an area to improve the stage distribution of diagnosed cases of cancer of the cervix, in the hope of reducing morbidity from the disease and potentially reducing mortality also. Downstaging should be directed to the appropriate age group to ensure cost-effective use of resources (women age 35 or preferably 45 and more). Experience in pilot projects in India has already shown that there are a number of cultural barriers to acceptance of downstaging, from the woman herself, from her family, and even from health care workers. Further, even if women are examined and found to have suspect abnormalities, referral to a district hospital for diagnosis and therapy can encounter great barriers. It is therefore important that if it is decided to introduce such a program in an area, appropriate arrangements are made to evaluate the effectiveness of the program.

Another cancer for which downstaging should be considered where incidence is high is oral cancer. One of the ten most common cancers in the world, oral cancer accounts for about a third of all cancers in Bangladesh, India, Pakistan, and Sri Lanka. The possibility of controlling oral cancer through early diagnosis in such countries appears to be good. First, the oral cavity is accessible for routine examination. Second, nonmedical personnel can readily detect lesions that are precursors of carcinoma [6]. Finally, there are indications that these precursor lesions may regress with cessation of tobacco exposure; and surgical treatment of early oral cancer is very effective. Experience in Southeast Asia has demonstrated under field conditions that primary health care workers can examine large numbers of people and can detect and classify precancerous lesions and cancers of the oral region with acceptable accuracy.

Thus downstaging for mouth cancer should encourage programs of "look a friend in the mouth" or self-examination using a mirror. If resources permit, encourage allied health workers to perform routine oral examination of adult smokers and chewers of tobacco.

For breast cancer, downstaging may be accomplished by breast self-examination and/or by physical examination by a health care professional ([7]; see also later section of this chapter).

For skin cancer, examination for tropical ulcers and appropriate therapy are important in many countries, while in other countries surveillance for the early signs of melanoma can be advocated, especially among those with fair skin (see chapter 10, this volume).

In each instance, it is important that measures are introduced to ensure that those suspected of having cancer are promptly referred for appropriate diagnosis and therapy, and that institutions are identified with the appropriate staffing and facilities to provide effective treatment and with accessibility to

the patients. Special measures may have to be introduced to ensure that those referred do attend for diagnosis and management of detected abnormalities.

3. Screening for cancer in developing countries

3.1. General principles

The principles that apply to the introduction of cancer screening in developing countries are identical to those accepted for developed countries, namely,

- the type of disease should be common and carry high morbidity and/or mortality;
- effective treatment known to reduce morbidity and mortality from the disease following screening should be available; and
- the screening procedure should be acceptable to people, relatively inexpensive, and safe.

Screening concentrating on high-risk groups is rarely justified on its own, since identified risk groups usually account for only a small proportion of the cancer burden in a country. However, in planning recruitment into screening programs, measures must be introduced to ensure that those at high risk are included.

3.2. Screening for cancer of the cervix

Cervical cancer is the most common cancer among women in developing countries and the second most common worldwide, with half a million new cases each year [1]. Screening with the cervical smear plus adequate follow-up therapy can achieve a major reduction in both incidence and mortality rates for cervix cancer [8].

Quantitative studies have shown that, after one negative cytologic test for cervix cancer, screening once every 3 to 5 years accomplishes about the same effect among women 35–64 years of age as screening every year (see chapter 4, this volume). Even screening once every 10 years yields almost two thirds reduction in frequency of invasive cervical cancer [9]. This evidence led a WHO meeting to conclude that in countries where resources are limited, the aim should be to screen every woman once in her lifetime between 35 and 40 years of age. When more resources are available, the frequency of screening should be increased to once every 10 and then every five years for the age groups 35 to 55 years and, ideally, once every three years for women aged 25 to 60 years [10]. In developing countries, the aim should be to screen every woman age 35–40 once in her lifetime if laboratories are available to examine the smears and facilities are available for treatment of abnormalities. Once 80% of women age 35–40 have been screened once, the frequency should be increased to 10- and then 5-year screening over the age range 30–60 as resources permit.

The extent to which the evidence on effectiveness of breast cancer screening (chapter 6, this volume) can be extended to developing countries needs evaluation. Unfortunately, mammography is an expensive test that requires great care in its delivery and expert attention to quality control in performing and reading the test [11]. It is therefore currently out of reach for many developing countries [12].

Breast Self-Examination (BSE) has the potential to improve the outlook for interval cancers, while its teaching has probably helped diminish false reassurance in those studies where it has been advocated together with other screening tests. The World Health Organization has concluded that only BSE has the potential to provide early diagnosis of breast cancer in many parts of the world [13]. Three case-control studies evaluating the possible role of BSE in reducing advanced breast cancer have been conducted. The first two were reported as negative [14,15], though in one [15] there appeared to be some benefit in compliers with BSE. This latter finding has now been confirmed in the Canadian National Breast Screening Study (see chapter 6, this volume). An important finding was the apparent similarity of effect in women under and over the age of 50. A cohort study has also been conducted of the Mama BSE program introduced by Gastrin in Finland in 1973. The records of nearly 30,000 women who enrolled in the program from 1973 to 1975 and who returned BSE calendars to Dr Gastrin were identified, and passive follow-up was conducted by record linkage to the Finnish Cancer Registry. These BSE compliers had elevated breast cancer incidence but reduced breast cancer mortality compared to that expected from the Finnish population [16]. Again, there did not seem to be a differential effect in women under and over the age of 50. The other two studies are randomized trials. The first, which is being conducted in Russia with group randomization by factory in Moscow and polyclinic in St Petersburg, has enrolled nearly 200,000 women, half of whom have been instructed in BSE. Follow-up is expected to last through to 1998 at least [17]. The other trial, planned to enroll about 300,000, is underway in Shanghai.

An indication that good physical examinations of the breast performed by specially trained health workers could have an important role has come from the Canadian National Breast Screening Study, where the addition of mammography to such examinations in women age 50–59 did not result in reduced breast cancer mortality [18]. This finding, coupled with the fact that the first trial of breast cancer screening, the Health Insurance Plan Trial of Greater New York, showed a breast cancer mortality reduction even with very early mammography [19], with the possibility that up 70% of the benefit came from the physical examination component of the screen [20], suggests that screening by physical examination of the breasts may have been undervalued.

Thus in most countries, the aim for breast cancer screening should be to

encourage BSE, and if possible participation in evaluation of physical examination for women age 40–69. If mammography becomes available in a country, the first priority for its use is for diagnosis, especially for women who have detected an abnormality on BSE, though cancer may be present even if the mammogram is negative. Introduction of mammography for screening should be avoided unless or until resources are available to ensure high-quality services sufficient to screen at least 70% of the target age group (on current evidence, only women age 50–69) [21].

3.4. Screening for other sites of cancer

As reviewed in other chapters in this volume, screening for many cancer sites must be regarded as experimental and cannot be recommended at present as public health policy. Screening should only be introduced in developing countries where the relevant cancer is a major problem and when planned as a research or demonstration project with mechanisms to evaluate the effectiveness of the programe. This includes screening for cancer of the esophagus, gastric cancer, colon and rectal cancer, liver cancer, lung cancer, ovarian cancer, bladder cancer, and prostate cancer.

4. Conclusion

There is a temptation to assume that recommendations appropriate for developed countries on screening for cancer can be translated to developing countries with little modification. This assumption is false. The health care priorities of developing countries are very different. The first priority in cancer control must be to attempt to reduce the impact of the emerging epidemic of tobacco-associated cancers. For other cancers with evidence on the effectiveness of screening in developed countries, namely, breast over the age of 50 and cervix, the principles of National Cancer Control Planning [2] must be adhered to — primarily the need to ensure, after a situation analysis and review of public and professional knowledge and facilities for screening and treatment, that screening is a sufficient priority for action. It will then often be concluded that the first priority is to concentrate on public and professional education to facilitate early detection, and only then to introduce screening in the most cost-effective way possible.

References

1. Parkin DM, Pisani P, Ferlay J. 1993. Estimates of the worldwide incidence of eighteen major cancers in 1985. Int J Cancer 54:594–606.
2. World Health Organization. 1995. National Cancer Control Programmes. Policies and Managerial Guidelines. Geneva: World Health Organization.
3. Stjernswärd J. 1990. National training of radiotherapists in Sri Lanka and Zimbabwe: priori-

ties and strategies for cancer control in developing countries. Int J Radiat Oncol Biol Phys 19:1275–1278.
4. Ponten J, Adami HO, Bergstrom R, Dillner J, Friberg LG, Gustafsson L, Miller AB, Parkin DM, Sparen P, Trichopoulos D. 1995. Strategies for global control of cervical cancer. Int J Cancer 60:1–26.
5. Stjernswärd J, Eddy D, Luthra U, Stanley K. 1987. Plotting a new course for cervical cancer screening in developing countries. World Health Forum 8:42–45.
6. World Health Organization. 1984. Control of oral cancer in developing countries. Bull WHO 62:817–830.
7. Bassett AA. 1985. Physical examination of the breast and breast self-examination. In Miller AB (ed.), Screening for Cancer. Orlando, FL, Academic Press, pp. 271–291.
8. Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC. 1985. Evaluation of screening programmes for gynaecological cancer. Br J Cancer 52:669–673.
9. IARC Working Group. 1986. Summary chapter. In Hakama M, Miller AB, Day NE (eds.), Screening for Cancer of the Uterine Cervix (IARC Scientific Publications no. 76). Lyon: International Agency for Research on Cancer, pp. 133–144.
10. World Health Organization. 1986. Control of cancer of the cervix uteri. Bull WHO 64:607–618.
11. Miller AB, Tsechkovski M. 1987. Imaging technologies in breast cancer control: summary of a report of a World Health Organization meeting. Am J Roentgenol 148:1093–1094.
12. Miller AB. 1989. Mammography: a critical evaluation of its role in breast cancer screening, especially in developing countries. J Public Health Policy 10:486–498.
13. Miller AB, Chamberlain J, Tsechovski M. 1985. Self-examination in the early detection of breast cancer. A review of the evidence, with recommendations for further research. J Chron Dis 38:527–540.
14. Newcomb PA, Weiss NS, Storer BE, et al. 1991. Breast self-examination in relation to occurrence of advanced breast cancer. J Natl Cancer Inst 83:260–265.
15. Muscat JE, Huncharek MS. 1991. Breast self-examination and extent of disease: a population-based study. Cancer Detect Prev 15:155–159.
16. Gastrin G, Miller AB, To T, et al. 1994. Mortality from breast cancer in the Mama program for breast screening in Finland, 1973 to 1986. Cancer 73:2168–2174.
17. Semiglazov VF, Sagaidack VN, Moiseyenko VM, et al. 1994. Study of the role of breast self-examination in the reduction of mortality from breast cancer. Eur J Cancer 29A:2039–2046.
18. Miller AB, Baines CJ, To T, et al. 1992. The Canadian National Breast Screening Study: breast cancer detection and mortality in women age 50–59 on entry. Can Med Assoc J 147:1477–1488.
19. Shapiro S, Venet W, Strax P, Venet L. 1988. Periodic Screening for Breast Cancer. The Health Insurance Plan Project and its Sequelae, 1963–1986. Baltimore: The Johns Hopkins University Press.
20. Miller AB. 1991. Is routine mammography screening appropriate for women 40–49 years of age? Am J Prev Med 7:55–62.
21. Miller AB, Chamberlain J, Day NE, Hakama M, Prorok PC. 1990. Report of a workshop of the UICC project on evaluation of screening for cancer. Int J Cancer 46:761–769.

Index